ST. PETERSBURG INSTITUTE
FOR INFORMATICS AND AUTOMATION
OF THE RUSSIAN ACADEMY OF
SCIENCES (SPIIRAS)

AIR FORCE OFFICE OF SCIENTIFIC
RESEARCH (AFMC)
EUROPEAN OFFICE OF AEROSPACE
RESEARCH AND DEVELOPMENT

Applied Methods and Models of Knowledge Engineering in Information Based Health Assessment Systems

Final Report

Contract No.F61775-98-WE116

Contractor:
Chief Scientist
of the St. Petersburg Institute for
Informatics and Automation
Ph.D. Professor
V.I. Gorodetski

St. Petersburg, Russia
1999

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE July 1999	3. REPORT TYPE AND DATES COVERED Final Report

4. TITLE AND SUBTITLE

Applied Methods and Models of Knowledge Engineering in Information Based Health Assessment Systems.

5. FUNDING NUMBERS

F61775-98-WE116

6. AUTHOR(S)

Dr. Vladimir Gorodetski

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

St.Petersburg Institute For Informatics & Automation of the Russian Academy of Sciences
39, 14th Liniya
St. Peterburg 199178
Russia

8. PERFORMING ORGANIZATION REPORT NUMBER

N/A

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

EOARD
PSC 802 BOX 14
FPO 09499-0200

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

SPC 98-4066

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

A

13. ABSTRACT (Maximum 200 words)

This report results from a contract tasking St.Petersburg Institute For Informatics & Automation of the Russian Academy of Sciences as follows: The contractor will investigate knowledge engineering as applied to system diagnostics by rigorous mathematical justification for already existing diagnostic procedures utilizing statistical clustering and Bayes' technique. A software package implementing cluster analysis of statistical data, including selection and evaluation of particular subspaces in the factor space, and facilitating definition of cluster models will be developed. The software will utilize computer graphics to facilitate previewing of the clustering pattern in particular subspaces. Applications of Algebraic Bayes' Network approach to various classes of applications of experts' analysis techniques to diagnostics related problems will be considered. Further, a software tool and numerical examples utilizing model data, demonstrating the concept, will be developed. In addition, applications of classical regression analysis to such problems of diagnostics as estimation of remaining life expectancy of hardware subjected to adverse effects, and analysis of combined effects of several adverse conditions of hardware failure will be subjected to analysis, and implemented in software. A software tool and numerical examples utilizing model data, demonstrating the concept, will be developed.

14. SUBJECT TERMS

EOARD, Modelling & Simulation, Microelectronics, Diagnostics

15. NUMBER OF PAGES

91

16. PRICE CODE

N/A

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19, SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

Contents

Preface

The present volume is Final Report supposed to be submitted by schedule of works in accordance with the contract No. F61775 98-WE116. The Report describes all accomplishments, results and conclusions of the research for contract. In particular, it contains results of research on the following tasks:

- Mathematical analysis of existing diagnostic procedure utilizing statistical clustering and Bayes' techniques and further development of mathematical basis and algorithm for diagnostic model design.

As a whole, these results formed a new technology of assessment and prognosis of the probability of failure of the hardware like avionics;

- Development of software package implementing all steps of cluster analysis of statistical data and further phases of the developed technology.

This software made it possible to verify and validate the developed technology numerically. A peculiarity of the developed software is (1) the interactivity, (2) the utilization a computer graphics to facilitate previewing of the clustering pattern in particular subspaces and (3) the use of the modern software engineering environment like Visual C++ 5.0 and MS Access-97;

- Application of Algebraic Bayes' Network approach to various classes of applications of experts' analysis techniques to diagnostic related problem;
- Application of classical regression analysis to estimation of remaining life expectancy (residual performance resource) of hardware subjected to adverse effects.

Advantage of the developed regression model is that it is based on Dynamic Data Model (DDM) what has made it possible to design the regression model utilizing ideas of prognosis of time series.

All results provided by contract are presented in the sufficient details in the Final Report. Nevertheless, the Interim Report [IR-98] has to be considered as the indefeasible part of the former report because some aspects of the developed knowledge engineering technology for diagnostic related problem solving are considered in the latter in more details. In addition, Interim Report presented more numerical results which were not repeated in Final Report.

Final Report consists of six sections.

The *Section 1* is introductory. It outlines the application itself and related prognostic tasks, main focus of the research and contents of the Report.

Section 2 presents Dynamic Data Model that was developed as alternative to Static Data Model which was used for numerical validation of the developed prognostics model in the first phase of research resulting in Interim Report [IR-98].

Section 3 presents developed variants of the regression models and their comparative numerical analysis.

Section 4 describes main results related to the task of knowledge discovery from statistical database to design information-based health assessment system. This question is a focus of the research.

Section 5 is devoted to the application of Algebraic Bayes' Network approach to the diagnostic related knowledge engineering tasks. Theoretical part of the section coincides mainly with the respective material given in the Interim Report [IR-98]. In contrast, it contains much more numerical results which were calculated on the basis of the developed software.

Section 6 has to be considered as the general conclusion of the research. It outlines the main results, research contribution and summarizes the most perspective future research in the framework of prognostics and related topics of Knowledge Discovery from Data (KDD) technology.

This contents reflects the tasks formulated within contract.

Contractor
Chief Scientist of the St. Petersburg Institute for Informatics
and Automation of the Russian Academy of Sciences
Ph.D. Prof.

Vladimir Gorodetski.

1. Introduction

Safe, reliable and efficient operation of avionics is crucial for a modern aircraft or spacecraft. While in operation, avionics components are exposed to electrical perturbations, mechanical vibrations, excessive temperatures, humidity, etc. These adverse conditions, acting individually and in combination, are known to have cumulative effects leading to avionics performance degradation and failures. Until recently, it was virtually impossible to obtain data characterizing performance of individual units. At the present, availability of dedicated monitoring systems and like devices allows for the collection of large amounts of actual data of any particular unit of aircraft hardware. Based on this data, modern Data Mining techniques, common in technology of Knowledge Discovery from Data (KDD), made it possible to facilitate formulation and solution of important on-line and off-line prognostic-related problems.

These new possibilities for hardware monitoring, for on-line and off-line prognostic related problem solving predetermined the tasks that are the subject of the contract. According to the contract the research presented in this Report aimed at the development of mathematical models, algorithms and software for solving the following tasks:

- accurate assessment of the probability of failure of hardware, such as avionics, on the basis of its known *history of abuse* by environmental and operational factors;
- prognosis of the probability of failure of hardware at a given time in the future, for example, at the end of the forthcoming sortie of the aircraft;
- accurate assessment of the residual performance resource of hardware on the basis of regression model and its known *history of abuse* by environmental and operational factors and known cumulative time of maintenance (number of sortie).

These task statement was prompted by the modern concept of maintenance known as the "service when needed" [Skormin et al-97]. Let us consider the peculiarity of the above task statement compared to the traditional one.

Traditionally, reliability of any technical device (electronic, electro–mechanical, and mechanical) is defined in terms of such characteristics as the average time of normal (no-failure) operation. These reliability concepts referring to a statistically–generic device may be considered acceptable as long as the failures are caused by the factors related to manufacturing. At the present, this approach is not always acceptable. Manufacturers of electronics, due to completely automated processes, have achieved a very high degree of reliability of their products and very little variation in properties from device to device. Manufacturing-related effects on failures of electronics are gradually becoming less significant. The main causes of failures are traced now to the individual operational and environmental conditions of particular units. Therefore, the average time of normal operation and other "traditional" reliability characteristics, defined without taking into account actual "history of abuse" of a device, are becoming less important.

Classical reliability had a good reason for addressing a statistically-generic device. At that time it was virtually impossible to obtain data characterizing performance of individual units in various operating environments. At the present, availability of Time-Stress Measurement Devices (TSMD) [Popyack-98], smart sensors and data acquisition systems makes possible to collect large amount of actual data of any particular unit of aircraft hardware. Based on this data, modern Data Mining techniques, common in Knowledge Discovery from Data (KDD) technology, facilitate formulation and solution of important on-line and off-line reliability-related problems. The most important problem is forecasting the probability of failure of flight-critical units of aircraft hardware during a forthcoming sortie. Solving such problem implies the investigation of the role of various environmental factors in the development of particular failures, investigation of combined effects of several factors, reevaluation of probability of failure on the basis of known exposure to particular adverse conditions, as well as development of special types of mathematical models and model-based techniques.

Data Mining and KDD address specific practical needs for solving above-mentioned problems. Data Mining provides a wide spectrum of available techniques and tools to develop a KDD technology focusing on design of a mathematical model for particular application ([Frawley et al -91], [Matheus et al 93], [Fayyad et al-95-1], [Fayyad et al 95-2], [Bradley et al-98-1]). It is well known that every particular application possesses specific properties that require either the ability to adapt already existing Data Mining techniques or develop new ones to build an adequate and efficient technology of original data processing aimed at a particular model development.

As per common understanding [Fayyad et al-95-1], a KDD process considered herein consists of a number of Data Mining procedures that, regardless of domain and particular task, conceptually, include such steps as (1) definition of the goal of the task, (2) collection or model-based generation of adequate statistical data and its preprocessing, (3) data reduction, transformation to find useful data patterns and its specifications and representations, including visualization if possible, (4) development of a KDD strategy that, actually, corresponds to the outline of the future technology as a number of steps of Data Mining, (5) selection, adaptation or development of Data Mining methods and algorithms intended for the realization of the accepted KDD technology (search of informative subsets of attributes and pattern of interest, separation and decision making rules creation, features, regression model development, etc.), (6) interpretation of the Data Mining results and incorporation of these results into a target model, (7) testing and validation of the resultant model.

Steps of this KDD process are usually iterative and interactive and are common for any KDD process. Nevertheless, from the algorithmic and implementation points of view, particular KDD processes may be implemented in very different ways. It is well known that the best universal approach does not exist. Moreover, the wider the area of possible utilization of an algorithm or approach, the lesser its efficiency. Therefore, taking into account the domain and task specifics, combined with the experience in KDD technology and Data Mining, assures the successful solution of any particular application problem. Then, following such a principle in the framework of tasks predetermined by contract, we developed an approach that consists of traditional steps of KDD process but its application reflects the following framework:

- peculiarities of the goal of the task (prognosis of probability of failure of avionics);
- original statistical data available for diagnostic and prognostic model design (for example, TSMD-based records of cumulative exposure to environmental factors and operational conditions);
- the need for a highly dependable model-based prognostic procedure;
- requirement of a reliable assessment of probabilities involved in the calculation of the probability of failure of a hardware even if the size of statistical data is small;

The Report is organized as follows.

In the first phase of research reflected in the Interim Report [IR-98] we considered «history of abuse» specified by the vector of adverse exposures of a unit operation and environmental conditions. This data model may be called "Static Data Model" (SDM), because it doesn't take into account the history of failure development. It was reasonable to use such simplified data model to focus research on the mathematical aspects regarding the task of prognostics.

Unfortunately, SDM is not appropriate for development of the precise regression model aimed at residual performance resource forecasting. It will be justified below that to solve the last task we need a model that reflects *the history of failure development for a particular device*, i.e. we need a model that makes it possible to specify the «trajectory» of failure development for a particular unit. It is reasonable to call the model that makes it possible to obtain trajectory of failure development of a particular unit as "Dynamic Data Model" (DDM). In the next section (*Section 2)* we present the developed Dynamic Data Model.

Section 3 is devoted to the presentation of the developed variants of the regression models for the forecasting of the residual performance resource of the hardware and its comparative numerical analysis. It was obtained numerically that traditional regression model that doesn't takes into account the history of a failure development of the particular device doesn't possess the required precision of residual performance resource forecasting. Instead, the regression model designed on the basis of DDM model of failure development seems to be much more advantageous. The corresponding regression model was developed and is described in *Section 3*. Additionally, this section contains results of numerical investigation of the regression procedure parameters that are sensitive regarding to the precision of the assessed residual performance resource.

In *Section 4,* to assess probability of failure of avionics, the developed technology of the model-based prognosis system is presented. Actually, these results were presented in detail in the Interim Report [IR-98]. Nevertheless, they outlined here in brief and are extended by some new numerical results obtained due to newly developed software. This section contains a brief description of the heuristic informativity criteria that are used in a general case for the preliminary selection of informative subspaces of low dimension. These procedures are demonstrated numerically for two-dimensional case.

A notion of a classification predicate is defined and a number of approaches to obtain such predicates are proposed. We use a visualization technique that makes it possible for a developer to draw a separation bound of any arbitrary form approximated by linear spline and to generate the associated classification predicate automatically. Then we describe the main principle behind the design of decision trees and associated probabilistic spaces that form a set of decision procedures. Since the major purpose of the model under development in this section is the assessment of the probability of failure of a hardware unit, in *Section 4* we consider the way of improvement of the precision and reliability of this assessment using the small size of experimental data and experts' knowledge. We present the numerical results as an example of an implementation of the outlined technology for the development of a model-based prognostic procedure for a particular avionics module.

Section 5 is devoted to the application of Algebraic Bayes' Network approach to the diagnostic related knowledge engineering tasks. Theoretical part of the section coincides in main aspects with the respective material given in the Interim Report [IR-98]. In contrast, it contains much more numerical results that were calculated on the basis of additionally developed software.

Section 6 has to be considered as the general conclusion of the research. It outlines the main results, research contribution and summarizes the most perspective future research in the framework of prognostics and related topics of Knowledge Discovery from Data (KDD) technology. They may be considered as the topics of the proposals for eventual future research.

2.Dynamic Data Model of Failure Development

2.1. Dynamic Data Model vs. Static Data Model

According to the problem statement [IR-98] health assessment system of a device on the board of an aircraft aims at solving two major tasks:

- Evaluation of the residual performance resource, given "history of abuse", and
- Prognosis of the probability of failure at a given time in the future, for example, at the end of the forthcoming sortie.

In the first phase of research reflected in the Interim Report [IR-98] we have considered «history of abuse» that was specified by the vector of adverse exposures of a device operation and environmental conditions. This data model may be called as "Static Data Model" (SDM), because it doesn't take into account the history of failure development of a particular device. It was reasonable to use such simplified data model to focus research on the mathematical aspects regarding the task of prognostics.

Unfortunately, SDM is not appropriate for development of the precise regression model of residual performance resource forecasting. It will be justified below that to solve the last task we need a model that reflects the history of failure development for a particular device, i.e. we need a model that makes it possible to specify the "trajectory" of failure development for a particular device. The latter is understood as ordered sequence of pairs $<t_i, X(t_i)>$, where - t_i is the cumulative time of device performance, $X(t_i)$ - is the vector of adverse exposures at the time t_i. It is reasonable to call the model which enables to obtain trajectory of failure development of a particular device as "Dynamic Data Model" (DDM).

On the other hand, regression model for residual performance resource forecasting results in one more sensitive parameter that may be used for the failure prognostics. Therefore, utilization of DDM instead of SDM makes possible to extend vector of adverse exposures by one more component, for example, number of aircraft sortie. As a result DDM makes it possible to evaluate in more realistic way the mathematical basis of prognostic task solving developed on the first phase of research [IR-98] and to improve it if necessary. Below in this section we describe the developed DDM for information-based health assessment system which is intended to solve both tasks mentioned at the beginning of this section.

2.2. Properties of DDM and Assumptions

We aim at developing DDM that is provided by the following properties:

- correlation of components of the vector of adverse exposures $X(t)$ at a time t of maintenance of a device is known and specified by their correlation matrix $C_X(t,t)$ and standard deviations $\sigma_X(t) = [\sigma[x_1(t)], \sigma[x_2(t)], ..., \sigma[x_n(t)]]^T = [\sigma_1(t), \sigma_2(t), ..., \sigma_n(t)]^T$; these mathematical entities define covariance matrix $W_X(t_i,t_i) = = M[(X(t_i) - \overline{X}(t_i))(X(t_i) - \overline{X}(t_i))]^T$ and, hence, random values of adverse exposures accumulated during a sortie of aircraft;
- mathematical expectation of adverse exposures per a sortie of any aircraft is constant and denoted by $M[\Delta X] = \Delta\overline{X}$[1];

[1] This assumption was accepted for simplicity of data model implementation and doesn't influence on the generality of data model itself.

- mathematical expectations of accumulated adverse exposures $M[X(t)]=\overline{X}(t)$ depends on the time *t* of device maintenance reflecting cumulative character of adverse factors;
- individual biases of mathematical expectations of adverse exposures $\delta\overline{X}(t,r)$ are randomized for every particular device number *r, r=1,2,...,R* (this property aims at taking into account the specific of manufacturing of a device and its maintenance conditions on the board of the particular aircraft);
- mutual correlation of the vectors of adverse exposures assigned to the different values of time of maintenance is specified by the matrix of mutual covariance $W_X(t_i,t_j)=$ $M[(X(t_i)-\overline{X}(t_i))(X(t_j)-\overline{X}(t_j))]^T$ that is supposed to be computed numerically from a statistical data base ;
- Realization of the random event $Q\in$ *{«no failure», «failure»}={0,1)*[1] is defined according to the truth values of logical formulae $F\in\Im$ given over the linear terms $Y_s=a_{1s}x_1+a_{2s}x_2+...+a_{ns}x_n$, where component x_i, *i=1, 2,...,* *n* are the components of the vector of adverse exposures, $a_{1s},a_{2s},...a_{ns}$ - are real valued coefficients and *s* – is the index of the vector of adverse exposures.

Additional assumptions utilized within the developed DDM model are as follows:

- correlation matrix $C_X(t,t)$ and standard deviations $\sigma_X(t)$ = $[\sigma_1(t),\sigma_2(t),...,\sigma_n(t)]^T$ are independent on the number of aircraft sortie, i.e. $C_X(t,t)=C_X$ and $\sigma_X(t)=\sigma_X$; therefore, covariance matrix $W_X(t_i,t_i)=W_X$ is constant as well;
- at average, each sortie of particular aircraft has 2 hours long; this assumption makes it possible to use the number of sortie (denote it by symbol *k*) as an equivalent of time and to deal with discrete parameter *k* instead of continuous one *t*;
- number of device *r* may be identified as the aircraft number;
- distributions of random values elsewhere below are normal or uniform.

2.3. Numerical characteristics of DDM

Thus, DDM of failure development of each individual device (belonging to the aircraft) number *r* as it was introduced in the previous section is defined by the following data:

- covariance matrix W_X that can be calculated in standard way via correlation matrix C_X and vector of standard deviation σ_X;
- mathematical expectations of adverse exposures $M[X(t)]=\overline{X}(t)$;
- individual biases of mathematical expectations of adverse exposures $\delta\overline{X}(k,r)$;
- number of sortie *k* of the concrete aircraft *r* and
- set of logic formulae $F_s\in\Im$ that determinate the conditions corresponding to a realization of the random event *"failure"*.

Therefore, DDM may be used for generation of realizations of trajectories $X(k,r)$ of failure development of a device on the board of the aircraft number *r=1, 2,..., R*. Below the information about adverse exposures and numerical data needed to generate the realizations of trajectories $X(k,r)$ are given.

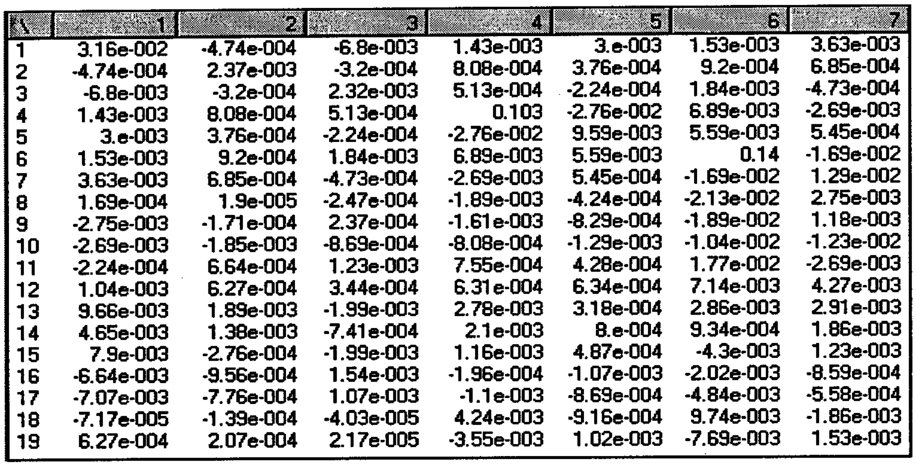
[1] We consider binary status of device performance within the designed DDM. Notice, that in the classification problem statement (see *Section 4*) we consider one more value of device status.

2.3.1. Table of adverse exposures (database composition)

X_1	Vibration RMS, 1 - 2 g,	X_6	Environmental Temperature 15 - 0°C	X_{11}	Power Supply 1.1 - - 1.3 nominal Vdc	X_{16}	Functional Overload 31 - 40% X_1
X_2	Vibration RMS, 3 - 4 g	X_7	Environmental Temperature 0 - 15°C	X_{12}	Power Supply over 1.3 nominal Vdc	X_{17}	Functional Overload 41 - 50%
X_3	Vibration RMS, over 4g	X_8	Environmental Temperature 50 - 75°C	X_{13}	Functional Over-load 5 - 10% X_1	X_{18}	Air Pressure .3 - .7 nominal
X_4	Humidity, 20 – 50%	X_9	Environmental Temperature 76 - 100°C	X_{14}	Functional Over-load 11 - 20% X_1	X_{19}	Air Pressure 1.1 - 1.3 nominal
X_5	Humidity, 70 – 95%	X_{10}	Power Supply .7 - .9 nominal Vdc	X_{15}	Functional Over-load 21 - 30%	X_{20}	Residual Performan-ce Resource (in hours)

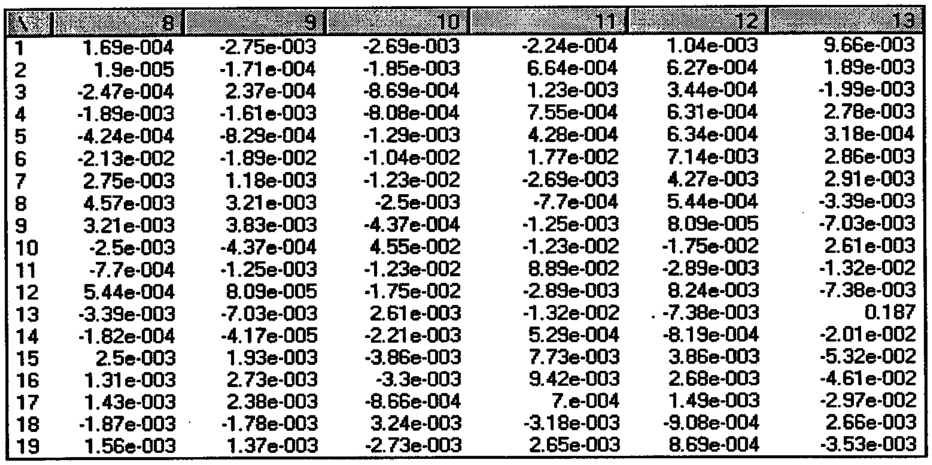
One more column (#21) of Database composition contains the value of device status and is omitted in the above table of Database composition.

2.3.2. Covariance Matrix W_X



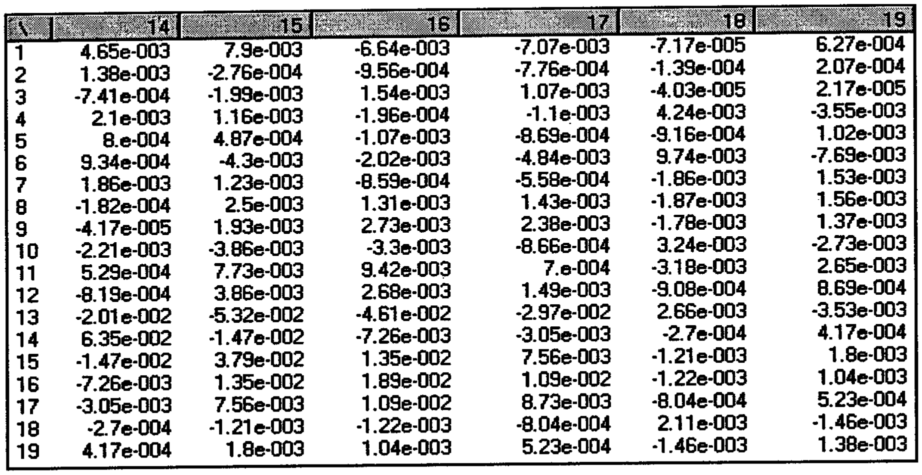
\	1	2	3	4	5	6	7
1	3.16e-002	-4.74e-004	-6.8e-003	1.43e-003	3.e-003	1.53e-003	3.63e-003
2	-4.74e-004	2.37e-003	-3.2e-004	8.08e-004	3.76e-004	9.2e-004	6.85e-004
3	-6.8e-003	-3.2e-004	2.32e-003	5.13e-004	-2.24e-004	1.84e-003	-4.73e-004
4	1.43e-003	8.08e-004	5.13e-004	0.103	-2.76e-002	6.89e-003	-2.69e-003
5	3.e-003	3.76e-004	-2.24e-004	-2.76e-002	9.59e-003	5.59e-003	5.45e-004
6	1.53e-003	9.2e-004	1.84e-003	6.89e-003	5.59e-003	0.14	-1.69e-002
7	3.63e-003	6.85e-004	-4.73e-004	-2.69e-003	5.45e-004	-1.69e-002	1.29e-002
8	1.69e-004	1.9e-005	-2.47e-004	-1.89e-003	-4.24e-004	-2.13e-002	2.75e-003
9	-2.75e-003	-1.71e-004	2.37e-004	-1.61e-003	-8.29e-004	-1.89e-002	1.18e-003
10	-2.69e-003	-1.85e-003	-8.69e-004	-8.08e-004	-1.29e-003	-1.04e-002	-1.23e-002
11	-2.24e-004	6.64e-004	1.23e-003	7.55e-004	4.28e-004	1.77e-002	-2.69e-003
12	1.04e-003	6.27e-004	3.44e-004	6.31e-004	6.34e-004	7.14e-003	4.27e-003
13	9.66e-003	1.89e-003	-1.99e-003	2.78e-003	3.18e-004	2.86e-003	2.91e-003
14	4.65e-003	1.38e-003	-7.41e-004	2.1e-003	8.e-004	9.34e-004	1.86e-003
15	7.9e-003	-2.76e-004	-1.99e-003	1.16e-003	4.87e-004	-4.3e-003	1.23e-003
16	-6.64e-003	-9.56e-004	1.54e-003	-1.96e-004	-1.07e-003	-2.02e-003	-8.59e-004
17	-7.07e-003	-7.76e-004	1.07e-003	-1.1e-003	-8.69e-004	-4.84e-003	-5.58e-004
18	-7.17e-005	-1.39e-004	-4.03e-005	4.24e-003	-9.16e-004	9.74e-003	-1.86e-003
19	6.27e-004	2.07e-004	2.17e-005	-3.55e-003	1.02e-003	-7.69e-003	1.53e-003

Covariance Matrix W_X (continuation)



\	8	9	10	11	12	13
1	1.69e-004	-2.75e-003	-2.69e-003	-2.24e-004	1.04e-003	9.66e-003
2	1.9e-005	-1.71e-004	-1.85e-003	6.64e-004	6.27e-004	1.89e-003
3	-2.47e-004	2.37e-004	-8.69e-004	1.23e-003	3.44e-004	-1.99e-003
4	-1.89e-003	-1.61e-003	-8.08e-004	7.55e-004	6.31e-004	2.78e-003
5	-4.24e-004	-8.29e-004	-1.29e-003	4.28e-004	6.34e-004	3.18e-004
6	-2.13e-002	-1.89e-002	-1.04e-002	1.77e-002	7.14e-003	2.86e-003
7	2.75e-003	1.18e-003	-1.23e-002	-2.69e-003	4.27e-003	2.91e-003
8	4.57e-003	3.21e-003	-2.5e-003	-7.7e-004	5.44e-004	-3.39e-003
9	3.21e-003	3.83e-003	-4.37e-004	-1.25e-003	8.09e-005	-7.03e-003
10	-2.5e-003	-4.37e-004	4.55e-002	-1.23e-002	-1.75e-002	2.61e-003
11	-7.7e-004	-1.25e-003	-1.23e-002	8.89e-002	-2.89e-003	-1.32e-002
12	5.44e-004	8.09e-005	-1.75e-002	-2.89e-003	8.24e-003	-7.38e-003
13	-3.39e-003	-7.03e-003	2.61e-003	-1.32e-002	-7.38e-003	0.187
14	-1.82e-004	-4.17e-005	-2.21e-003	5.29e-004	-8.19e-004	-2.01e-002
15	2.5e-003	1.93e-003	-3.86e-003	7.73e-003	3.86e-003	-5.32e-002
16	1.31e-003	2.73e-003	-3.3e-003	9.42e-003	2.68e-003	-4.61e-002
17	1.43e-003	2.38e-003	-8.66e-004	7.e-004	1.49e-003	-2.97e-002
18	-1.87e-003	-1.78e-003	3.24e-003	-3.18e-003	-9.08e-004	2.66e-003
19	1.56e-003	1.37e-003	-2.73e-003	2.65e-003	8.69e-004	-3.53e-003

Covariance Matrix W_X *(continuation)*



\	14	15	16	17	18	19
1	4.65e-003	7.9e-003	-6.64e-003	-7.07e-003	-7.17e-005	6.27e-004
2	1.38e-003	-2.76e-004	-9.56e-004	-7.76e-004	-1.39e-004	2.07e-004
3	-7.41e-004	-1.99e-003	1.54e-003	1.07e-003	-4.03e-005	2.17e-005
4	2.1e-003	1.16e-003	-1.96e-004	-1.1e-003	4.24e-003	-3.55e-003
5	8.e-004	4.87e-004	-1.07e-003	-8.69e-004	-9.16e-004	1.02e-003
6	9.34e-004	-4.3e-003	-2.02e-003	-4.84e-003	9.74e-003	-7.69e-003
7	1.86e-003	1.23e-003	-8.59e-004	-5.58e-004	-1.86e-003	1.53e-003
8	-1.82e-004	2.5e-003	1.31e-003	1.43e-003	-1.87e-003	1.56e-003
9	-4.17e-005	1.93e-003	2.73e-003	2.38e-003	-1.78e-003	1.37e-003
10	-2.21e-003	-3.86e-003	-3.3e-003	-8.66e-004	3.24e-003	-2.73e-003
11	5.29e-004	7.73e-003	9.42e-003	7.e-004	-3.18e-003	2.65e-003
12	-8.19e-004	3.86e-003	2.68e-003	1.49e-003	-9.08e-004	8.69e-004
13	-2.01e-002	-5.32e-002	-4.61e-002	-2.97e-002	2.66e-003	-3.53e-003
14	6.35e-002	-1.47e-002	-7.26e-003	-3.05e-003	-2.7e-004	4.17e-004
15	-1.47e-002	3.79e-002	1.35e-002	7.56e-003	-1.21e-003	1.8e-003
16	-7.26e-003	1.35e-002	1.89e-002	1.09e-002	-1.22e-003	1.04e-003
17	-3.05e-003	7.56e-003	1.09e-002	8.73e-003	-8.04e-004	5.23e-004
18	-2.7e-004	-1.21e-003	-1.22e-003	-8.04e-004	2.11e-003	-1.46e-003
19	4.17e-004	1.8e-003	1.04e-003	5.23e-004	-1.46e-003	1.38e-003

While designing the correlation matrix W_X we took into account the actually existing dependencies between adverse factors. These dependencies were extracted from the expert.

The correlation of adverse exposures $x_1 - x_{19}$, on the one hand, and residual performance resource x_{20}, on the other hand, have to be computed via simulation.

2.3.3. Vector of standard deviations σ_X

$\sigma(x_1)=0.18,\ \sigma(x_2)=0.049,\ \sigma(x_3)=0.048,\ \sigma(x_4)=0.321,\ \sigma(x_5)=0.097,\ \sigma(x_6)=0.04,$
$\sigma(x_7)=0.113,\ \sigma(x_8)=0.068,\ \sigma(x_9)=0.061,\ \sigma(x_{10})=0.067,$
$\sigma(x_{11})=0.198\ \sigma(x_{12})=0.09,\ \sigma(x_{13})=0.432,\ \sigma(x_{14})=0.252,\ \sigma(x_{15})=0.195,\ \sigma(x_{16})=0.137,$
$\sigma(x_{17})=0.934,\ \sigma(x_{18})=0.046,\ \sigma(x_{19})=0.037.$

According to the well known algorithm components of the covariance matrix W_X of the vector *X(k,r)* was calculated as follows:

$$W_X(x_i, x_j) = c_X(i,j) \times \sigma_i \sigma_j,$$

where $c_X(i,j)$ - are the elements of the correlation matrix C_X.

2.3.4.Predicate F: description of the status "failure"

$$F_1 = \{x_5 \geq 25\},\ F_2 = \{x_{12} \geq 27\}, F_3 = \{x_9 \geq 28\},\ F_4 = \{x_{19} \geq 27\},$$

$$F_5 = \{0.1 \times x_1 + 0.2 \times x_2 + 0.7 \times x_3 \geq 30\} = \{Y_1 \geq 30\},$$
$$F_6 = \{0.05 \times x_6 + 0.1 \times x_7 + 0.20 \times x_8 + 0.6 \times x_9 + 0.04 \times x_5 - 0.01 \times x_4 \geq 30\} = \{Y_2 \geq 30\},$$
$$F_7 = \{0.02 \times x_{10} + 0.3 \times x_{11} + 0.68 \times x_{12} \geq 30\} = \{Y_3 \geq 30\},$$
$$F_8 = \{0.02 \times x_{13} + 0.08 \times x_{14} + 0.1 \times x_{15} + 0.3 \times x_{16} + 0.5 \times x_{17} \geq 30\} = \{Y_4 \geq 30\},$$
$$Y_5 = \{0.3\, x_{18} + 0.7\, x_{19}\},$$
$$Y_6 = \{1.5 \times x_5 - 0.5 \times x_4\},$$

$$FZ1 = \{0.2 \times Y_1 + 0.25 \times Y_2 + 0.3 \times Y_3 + 0.2 \times Y_5 + 0.05 \times Y_6 \geq 20\} = \{Z_1 \geq 20\},$$
$$FZ2 = \{0.2 \times Y_2 + 0.2Y3 + 0.4 \times Y_4 + 0.1 \times Y_5 + 0.1 \times Y_6 \geq 15\} = \{Z_2 \geq 15\},$$

$$FZ3=\{0.15\times Y_1+0.25\times Y_3+0.4\times Y_4+0.1\times Y_5+0.1\times Y_6\geq 16\}=\{Z_3\geq 16\},$$

$$FF=\{0.2\times Z_1+0,4\times Z_2+0.4\times Z_3\geq 17\}.$$

The logic condition of the event *"failure"* is as follows:

$F=F_1 \vee F_2 \vee F_3 \vee F_4 \vee F_5 \vee F_6 \vee F_7 \vee F_8 \vee FZ1 \vee FZ2 \vee FZ3 \vee FF=$*"true"*.

Otherwise the status of the device is *"no failure"*.

2.4. Generation of trajectories of failure development

We suppose that increment of adverse exposures per sortie consists of two components. One of them is randomized bias $\delta X(k, r)=[\delta x_1(k,r),\delta x_2(k,r),...,\delta x_{19}(k,r)]$ that has individual distribution for each device *r* and depends on the number of sortie *k* as well, and the second one is a random value $\Delta X(k)$ which distribution is independent on the individual properties of aircraft. We suppose that both of them have equal matrices of correlation and individual values of standard deviations.

Since the components of these vectors of adverse factors are correlated, each point of trajectory within DDM is generated in a number of steps. They are as follows:

1. Transformation of the vector of adverse factors to the form with non-correlated components.

Since we supposed in DDM that components of the random vector of adverse exposures X are correlated we have to use special algorithm of generation of its realization [see, for example, [Fukunaga-72].

Let W_X be the covariance matrix of the vector $\Delta X(k)$ of increment of adverse exposures per sortie number *k* of aircraft, and let B, Λ be matrices of eigenvectors and eigenvalues of the covariance matrix W_X. Then

$$W=B^T \Lambda B$$

and $Y=B^T X$ is a vector which components are non-correlated and distributed normally with $\sigma(y_i)=\sqrt{\lambda_i}$, where λ_i - is *i-th* diagonal element of the matrix Λ.

Thus, in the first step of the algorithm it is necessary to calculate matrices B and Λ as well as $\sigma(y_i)=\sqrt{\lambda_i}$.

2.Generation of the trajectories of failure development in terms of vector Y.

```
For $r=1,2,...,R$
    For $i=1,2,...,19$
    $\Delta y_i(0,r)=3\sigma(y_i)+\xi$ ($\xi$ - is normal random value having M$[\xi]=0$ and
    $\sigma(\xi)=\sigma(y_i)$)
    $\delta y_i(0, r)=\alpha\ \Delta y_i(0,r)$ ($\alpha$ - is uniform random value, $\alpha\in[0, \overline{\alpha}]$, we use $\overline{\alpha}=0,1$]).
    $x_i(0,r)=0$ end i.
k=0.
    While $\neg F$ do $k=k+1$
      For $i=1,2,...,19$
      $\delta y_i(k,r)=\delta_i(0,r)+\Delta y_i(0,r)\times\beta$, where $\beta$ - is uniform random value, $\beta\in[0,0.1]$;
```

```
        $\Delta y_i(k,r) = \gamma(3\sigma_i + \xi)$ ($\xi$ - is normal random value having $M[\xi]$=0 and
        $\sigma(\xi)=\sigma(y_i)$, where $\gamma$ - is uniform random value, $\gamma \in [0,1-\alpha-\beta]$)
        $\Delta y_i(k,r) = \delta y_i(k,r) + \Delta y_i(k,r)$ end i.
    3.Inverse transformation of the vector $\Delta Y$ to the vector $\Delta X$:
        $\Delta X(k,r)=B\Delta Y(k,r)$
    4. Calculation of the vector of cumulative adverse exposures
        $X(k,r) = X(k-1,r) + \Delta X(k,r)$
      end k.{Result: trajectory of failure development for given device (aircraft) # r}
end r
```

2.5. Simulation of DDM

The above DDM model implemented within Visual C++ 5.0 and Access 97 Data Base environment was used to generate statistical dynamic data supposed to be used for numerical validation of the developed mathematical model and algorithms of health assessment system and regression model. In the fig.2.1 – fig.2.4 the trajectories of failure development for selected adverse exposures for 25 samples of a device (aircraft) are given. In these figures horizontal axis corresponds to the number of sortie and vertical one corresponds to the value of cumulative exposure of respective adverse factor. Each trajectory consists of 60-120 points. Let us remind that the average duration of a sortie is equal to 2 hours long.

Each trajectory (a case) corresponds to a triple $<k,r,X(k,r)>$ where k – is the number of sortie, r - is the number of device (aircraft), $X(k,r)$ – is the vector of cumulative adverse exposure at the end of the sortie number k. In *Appendix A1* the trajectories for more components of adverse exposures development among factors $x_1(k,r)$-$x_{19}(k,r)$ are given.

Each trajectory has a final point that corresponds to the status " failure" of the device. Since the number of sortie at this point is known and we supposed that each sortie has 2 hours long, we can map each point of trajectory by one more purposeful variable, i.e. by variable $\tau(k,r)$ that has the sense of residual performance resource. Indeed, if $k_f(r)$ is the number of sortie that corresponds to the event *"failure"* and Δt is the duration of each sortie then

$$\tau(k,r) = k_f(r) \times \Delta t - k(r) \times \Delta t = [k_f(r) - k(r)]\Delta t. \qquad (2.1)$$

Formula (2.1) makes it possible to map each point of all trajectories of failure development by the value of residual performance resource $\tau(k,r)$. On the one hand, this mapping extends the statistical data in the way that makes possible to design a regression model (see *Section 3*). On the other hand, this mapping makes it possible to obtain one more sensitive parameter for prognosis the probability of failure at a given time in the future. In the fig.2.5 – fig.2.8 we depicted the trajectories of failure development in terms of the adverse factors but used the horizontal axis marked by the variable $\tau(k,r)$. In *Appendix A1* such trajectories are given for more components of adverse exposures development.

2.6. Interpretation of Simulated Data

Representation of statistical data in the form of the multitude of trajectories of failure development gives a new insight on the data interpretation. Formally, all points of any trajectory where $k(r) < k_f(r)$ correspond to the device status *"no failure"*. Nevertheless, it is intuitively clear that the points that are proximate to the points $X(k_f,r)$ form the *"border-line"* class of device

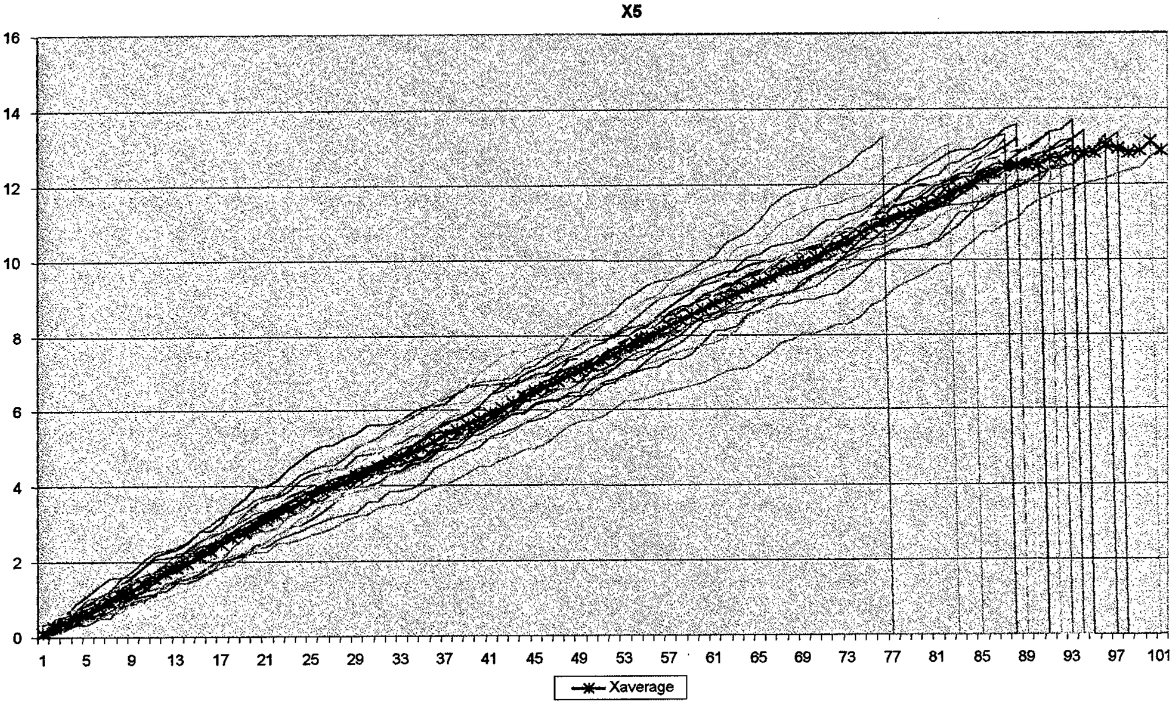


Fig.2.1. Realization of trajectories of development of adverse exposure X_5 as the functions of the number of aircraft sortie

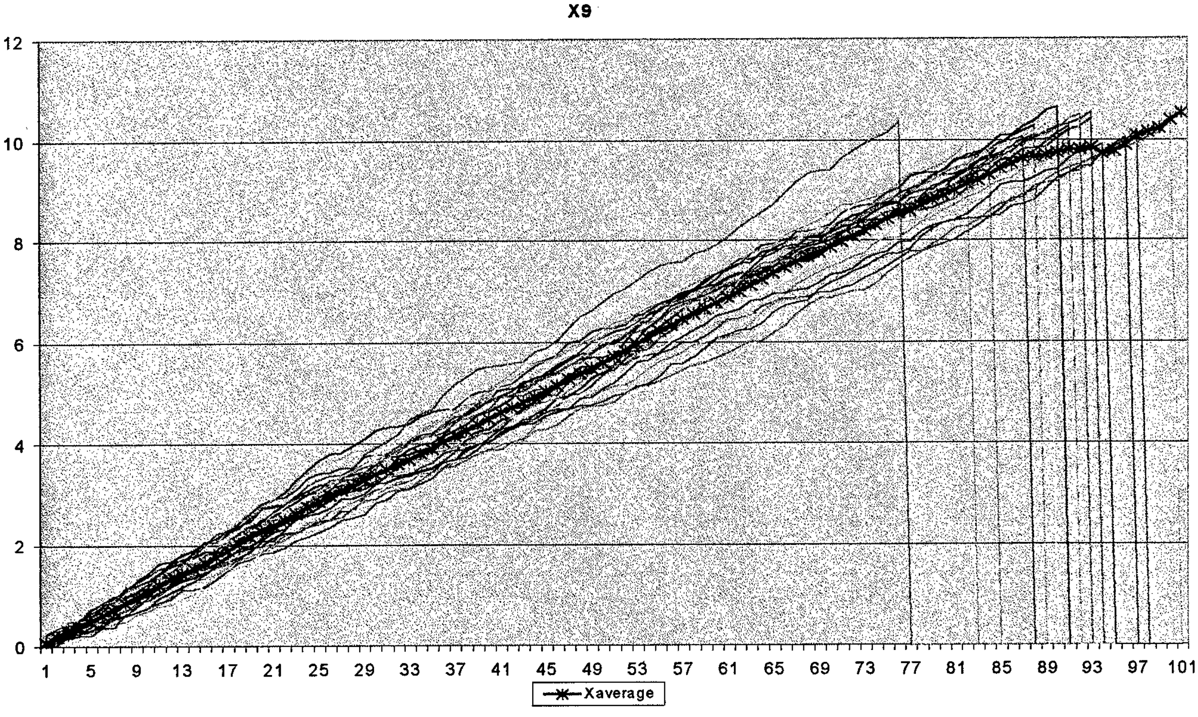


Fig.2.2. Realization of trajectories of development of adverse exposure X_9 as the functions of the number of aircraft sortie

2.Dynamic Data Model of Failure Development

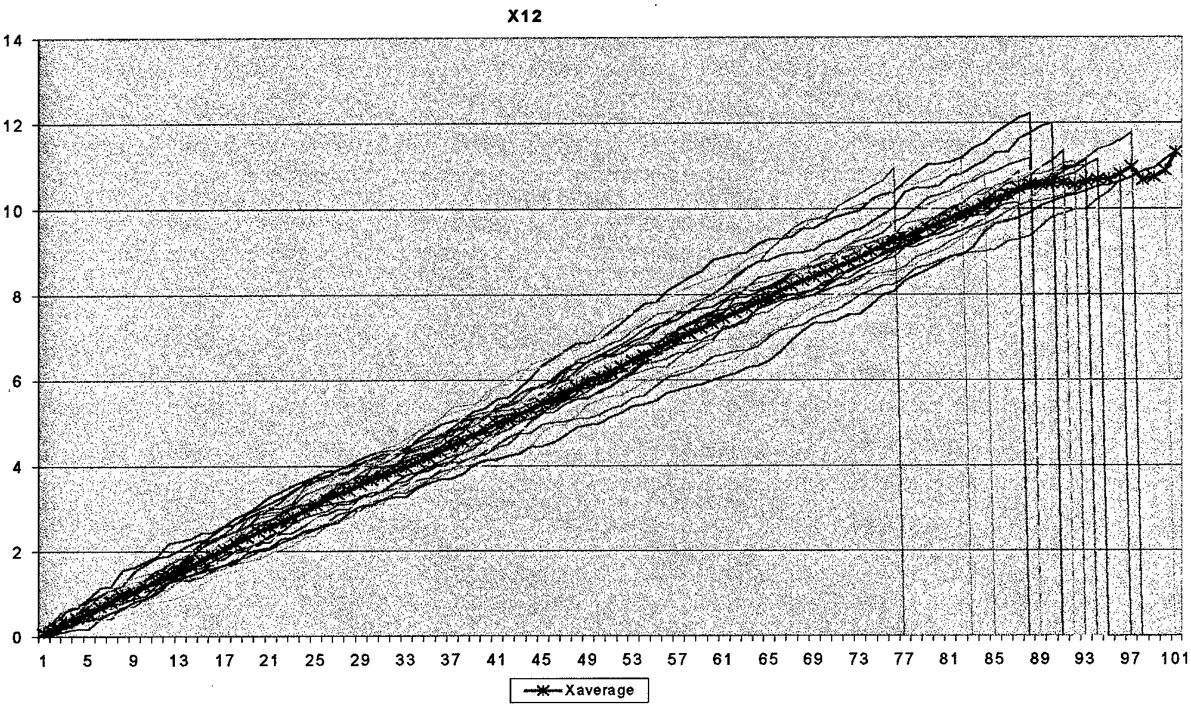


Fig.2.3. Realization of trajectories of development of adverse exposure X_{12} as the functions of the number of aircraft sortie

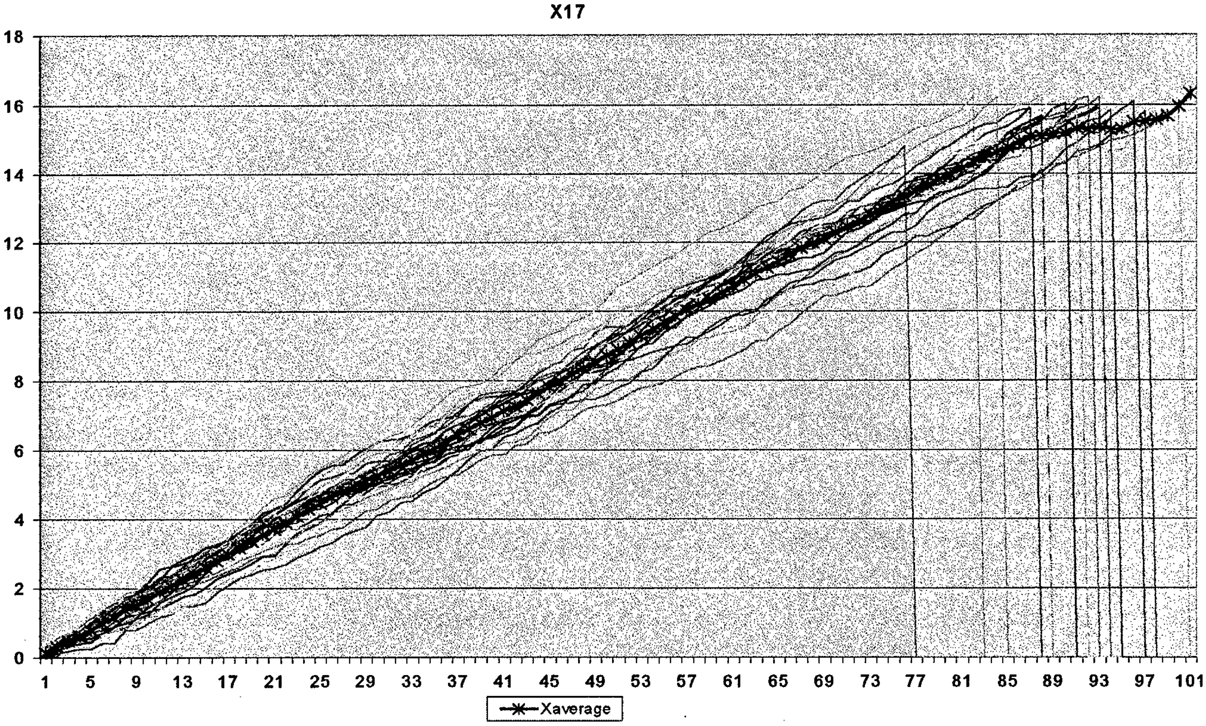


Fig.2.4. Realization of trajectories of development of adverse exposure X_{17} as the functions of the number of aircraft sortie

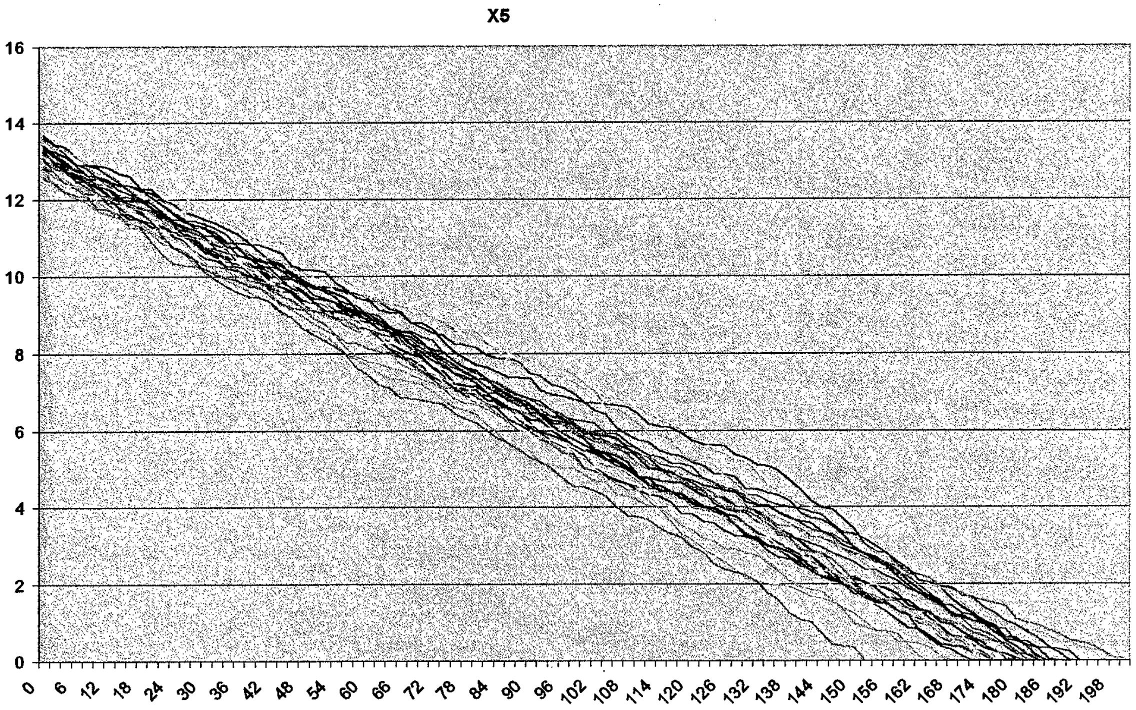


Fig.2.5. Realization of trajectories of development of adverse exposure X_5 as the functions of residual performance resource

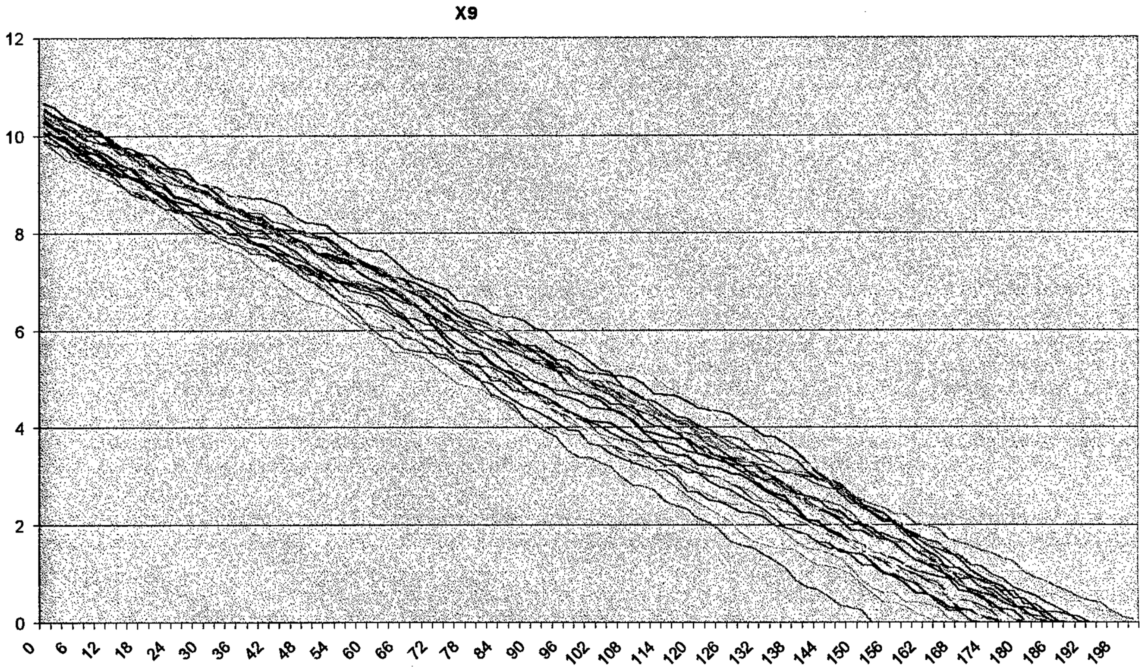


Fig.2.6. Realization of trajectories of development of adverse exposure X_9 as the functions of residual performance resource

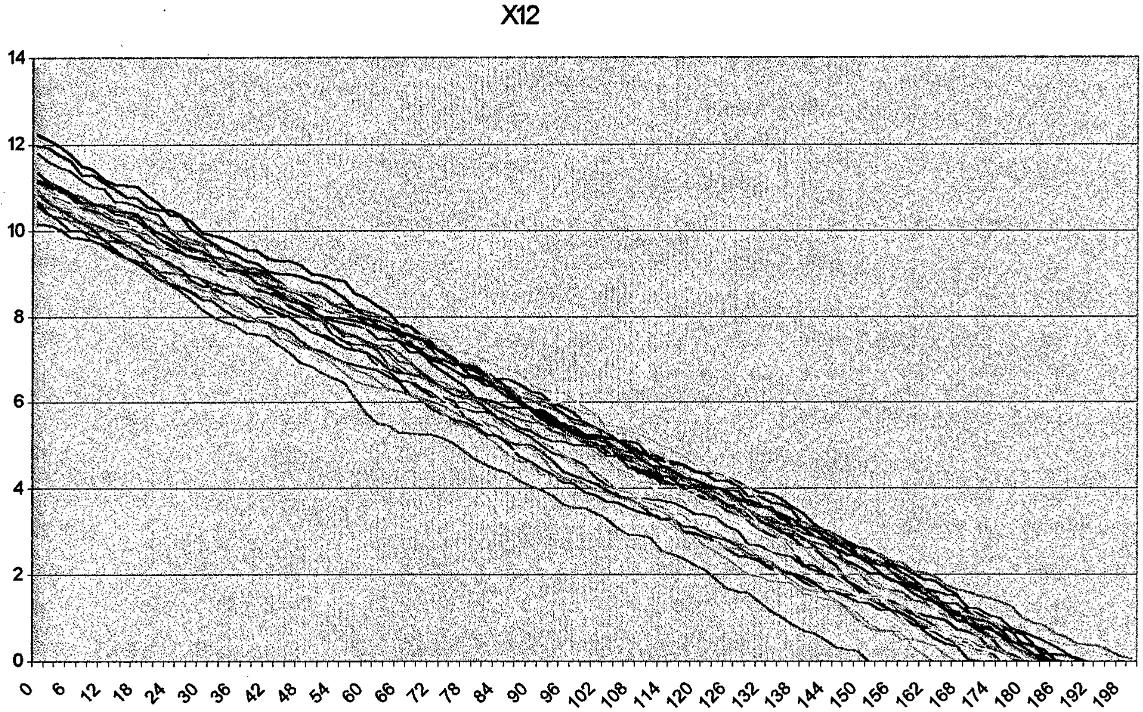


Fig.2.7. Realization of trajectories of development of adverse exposure X_{12} as the functions of residual performance resource

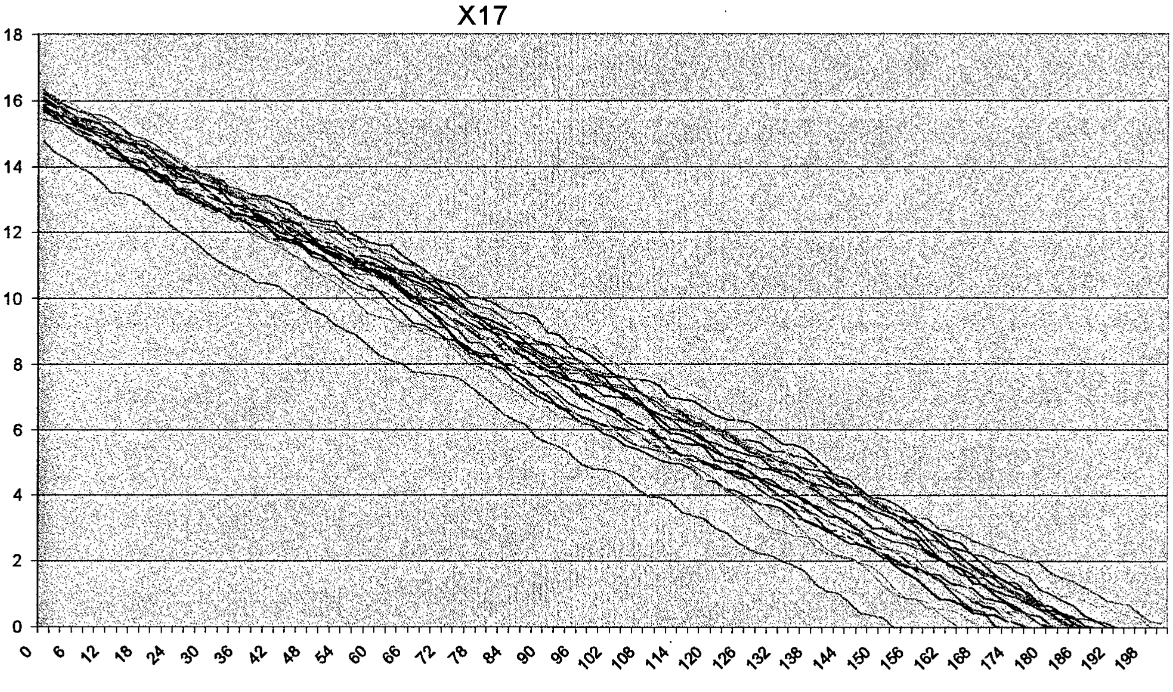


Fig.2.8. Realization of trajectories of development of adverse exposure X_{17} as the functions of residual performance resource

status which has to be learnt together with traditionally considered binary statuses *"no failure"-"failure"*.

This consideration is the reason of new database model used below for numerical validation of the developed mathematical model and algorithms of healthcare assessment system design. We used statistical database in which every case is assigned by a value of failure status from the set *{"no failure", "border-line", "failure"}={-1, 0, 1}*.

In *Appendix A2* the generated statistical database is given. It contains 200 cases *X(k,r)* each extended by the value of residual performance resource $\tau(k,r)$. These cases are used for knowledge engineering procedures to develop classification rules. 200 more cases of this database are used for testing of the resulting classification rules.

Let us note that cases included in database were chosen randomly among the points of all trajectories generated via DDM. As a result, we design a database of the same representation as one used in Interim Report [IR-98] The main distinction of these databases is that the latter used 3-valued interpretation of the status of devices performance. Let us consider this question in more details.

Together with conventionally used values of performance status *"no failure"*, *"failure"*, in this report we use one more value of status, i.e. the value *"border-line"*. This status is determined as follows:

$$\text{If } X \notin \{\text{"failure"}\} \text{ and } 0 < x_{20} \leq \bar{\tau} \text{ then } X \in \{\text{"border-line"}\}, \tag{2.2}$$

where *{"failure"}* - is the multitude of cases of database having status of performance *"failure"*; and *{"border-line"}* – is the multitude of all cases that have status of performance *"border-line"*; the value $\bar{\tau}$ – is a threshold of residual performance resource admissible to believe that status of device performance is "no failure". Below we accept the value of threshold as follows: $\bar{\tau} = 20$. This choice was conditioned by precision of regression model of the residual performance resource forecasting (see *Section 3*).

Numerical results obtained via DDM developed in this section turned out convenient to test the basic algorithms of the proposed prognostic model numerically (see *Section 4*).

3. Regression Model for Residual Performance Resource Assessment

3.1. Problem Statement and Traditional Approach

In this section we consider the task of regression model design for residual resource assessment. We suppose that initial information needed for design of the above model is given in the form of the statistical database of trajectories of failure development. The samples of such trajectories are depicted in the fig.2.4-fig.2.8 (Also see *Appendix A1.*).

The simplest formal problem statement of regression model design is as follows. It is given database of cases $<X, \tau>$, where X – is vector of cumulative exposures of adverse factors, τ - is the value of residual performance resource. The task is to design a function $\tau = f(X)$. Traditional regression model might be designed in the following way. Let us introduce the extended vector $\vec{X}^T = [X^T, \tau]$. While having database, we are able to calculate the vector of mathematical expectation and covariance matrix of the vector $\vec{X}^T = [X^T, \tau]$. Let us represent them as follows:

$$M[\vec{X}] = M[X^T, \tau]^T = [\overline{X}^T, \overline{\tau}]^T, \tag{3.1}$$

$$W[\vec{X}, \vec{X}] = \begin{bmatrix} W[X,X] & W[X,\tau] \\ W[\tau,X^T] & W[\tau,\tau] \end{bmatrix}, \tag{3.2}$$

where $\overline{X}$, $\overline{\tau}$ - are mathematical expectations of the respective variables and *W(*, *)* are covariance matrices of the variables within squared brackets

Dimensions of the blocks in matrix $W[\vec{X}, \vec{X}]$ correspond to the following scheme:

$$\begin{bmatrix} n \times n & n \times 1 \\ 1 \times n & 1 \times 1 \end{bmatrix} = \begin{bmatrix} 19 \times 19 & 19 \times 1 \\ 1 \times 19 & 1 \times 1 \end{bmatrix}.$$

In terms of accepted denotations the regression equation $\tau = f(X)$ is as follows:

$$\overline{\tau}(X) = M[\tau / X] = \overline{\tau} + W[\tau, X]W[X, X]^{-1}[X - \overline{X}]. \tag{3.3}$$

But this equation is not appropriate in practice because it doesn't depend on such a sensitive variable as number of sortie *k*. DDM described in the previous section makes it possible to design the more sophisticated and precise regression model.

3.2. DDM-based Regression Model

Let us consider database generated by DDM. For this database we are able to map each point of trajectory by the argument value "number of sortie", and to calculate the value of variable "residual performance resource". This means that we may design regression model in the following form ([Rao-71], [Anderson-60]):

$$\overline{\tau}(X,k) = M[\tau(k) / X(k)) = \overline{\tau}(k) + W[\tau(k), X(k)]W[X(k), X(k)]^{-1}[X(k) - \overline{X}(k)] \tag{3.4}$$

Comparing to the previous case (3.3), the peculiarity of the regression model (3.4) is that all statistics (mathematical expectations and matrices of covariance) are computed on the basis of sampling mapped by the same value of the number of sortie. Algorithm of their calculations are well known ([Rao-71], [Anderson-60]).

The next step of regression model improvement is the use of history of failure development and autoregression. Actually, for each sortie number *k* of an aircraft *r* the history *{X(1,r), X(2,r),..., X(k,r)}* is known. This history may be used to improve regression model in the following way.

Let us choose an integer value m. We say that regression model has the *depth of memory* equal to m if for any sortie number k the values of $X(k-m), X(k-m+1),...,X(k-1), X(k)$) are used for the regression model design. Let us show how the regression model of the depth m may be designed.

Let us introduce denotation $\vec{X}(k) =< X(k),\tau(k) >$ that is the vector of adverse factors extended by the value of residual performance resource at the end of a sortie of number k. Let us calculate the following covariance matrices associated with the vectors $\vec{X}(k-m)$, $\vec{X}(k-m+1),...,\vec{X}(k)$:

$$W(\vec{X}(k-m),\vec{X}(k-m)], W(\vec{X}(k-m),\vec{X}(k-m+1)],..., W(\vec{X}(k-m),\vec{X}(k)];$$

$$W(\vec{X}(k-m+1),\vec{X}(k-m+2)];,...,W(\vec{X}(k-m+1),\vec{X}(k)]$$

$$\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots$$

$$W(\vec{X}(k),\vec{X}(k)]$$

To reduce the above task to the standard form (3.4) of the regression model design let us compose the following block-wise matrices:

$$W[X(k,m)]=\begin{bmatrix} W[X(k-m),X(k-m)] & W[X(k-m),X(k-m+1)] & \dots & W[X(k-m)X(k)] \\ W[X(k-m+1),X(k-m)] & W[X(k-m+1),X(k-m+1)] & \dots & W[X(k-m+1),X(k)] \\ \dots & \dots & \dots & \dots \\ W[X(k),X(k-m)] & W[X(k),X(k-m+1)] & \dots & W[X(k),X(k)] \end{bmatrix} \quad (3.5)$$

$$W[k,m,X,\tau]=[W[X(k-m),\tau(k)] \;\; W[X(k-m+1),\tau(k) \;\; \dots \;\; W[X(k),\tau(k)]] \quad (3.6)$$

and

$$W[k,m,\tau]=\begin{bmatrix} W[X(k,m)] & W[k,m,X,\tau] \\ W^T[k,m,X,\tau] & \sigma^2[\tau(k)] \end{bmatrix} \quad (3.7)$$

As well let us suppose that all mathematical expectations used below are calculated.

Let we know the history of adverse exposures along the trajectory of failure development $X(k-m), X(k-m+1), ...,X(k)$. Let us denote this history as follows:

$$\mathbf{X}(k,m)=[X^T(k-m),X^T(k-m+1),...,X^T(k)]^T, \quad (3.8)$$

$$\overline{\mathbf{X}}(k,m) = M[\mathbf{X}(k,m)]. \quad (3.9)$$

While utilizing the formula like (3.4) for matrices $W[X(k,m)]$ (3.5), $W[k,m,X,\tau]$ (3.6) and mathematical expectation $\overline{\mathbf{X}}(k,m)$ (3.9), we can to constitute the following equation for assessment of the residual performance resource:

$$\overline{\tau}(k/\mathbf{X}(k,m))=\overline{\tau}(k)+W^T[k,m,X,\tau]W[X(k,m)]^{-1}[\mathbf{X}(k,m)-\overline{\mathbf{X}}(k,m)]. \quad (3.10)$$

3.3. Numerical results

We have investigated numerically what parameters of the regression procedure are sensitive regarding to the precision of the assessed residual performance resource. There were considered two of them, i.e. m – is the depth of memory, and s – is the interval between two points of trajectory (*"step"*) involved in regression model. The sense of the memory depth m has been explained already. The sense of the variable s is as follows. Let us use memory depth $m=2$. It means that to assess residual performance recourse we use three values of history of adverse exposures development, i.e.

$X(k_1)$, $X(k_2)$ and $X(k_3)$. Difference between two sequential values of variable *"number of sortie"* we mean as *step* of regression procedure, i.e. $s=k_2-k_1=k_3-k_2$.

It is clear that the more value of variable m the more computational complexity of regression procedure. Actually, increase of m entails remarkable increase of dimension of matrices in the equation (3.10). We investigated regression model precision for $m=0,\ 1,\ 2$ for different values of variable s. The results are given in the fig.3.1 –fig.3.8.

As a conclusion it was adjusted that the most appropriate value of m is equal to 2 and value of s is about (5–10). Decrease of s entails increase of noise but increase of it lead to decrease of precision. Of course, this conclusion is valid for concrete data model and sampling size for considered applied task.

The more important and traditionally discussed problem is what adverse factors have the most noticeable impact on the value of residual performance resource. Is it possible to diminish the size of the vector of adverse factors to minimum, for example, to 4-5 preserving the precision of the regression model within required limits? This task may be solved within the considered frameworks, but it takes to attract special approaches (not only statistical ones) and is a subject of special research.

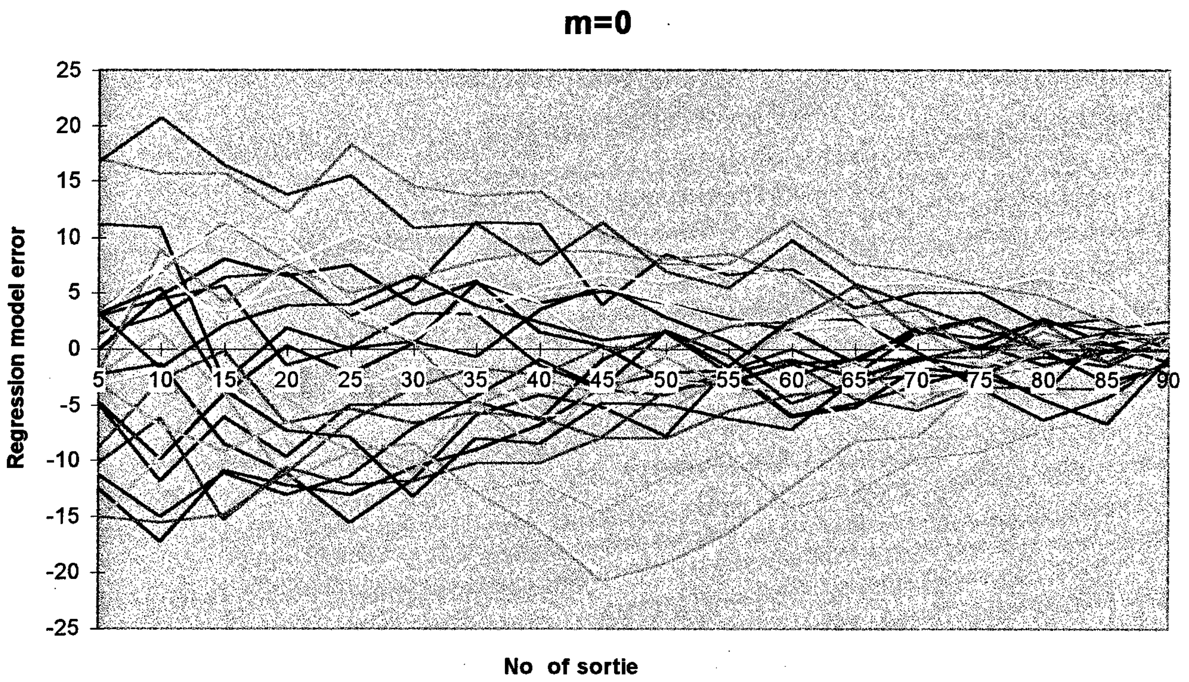


Fig.3.1. Realizations of error of residual performance résource assessment for *m=0.*

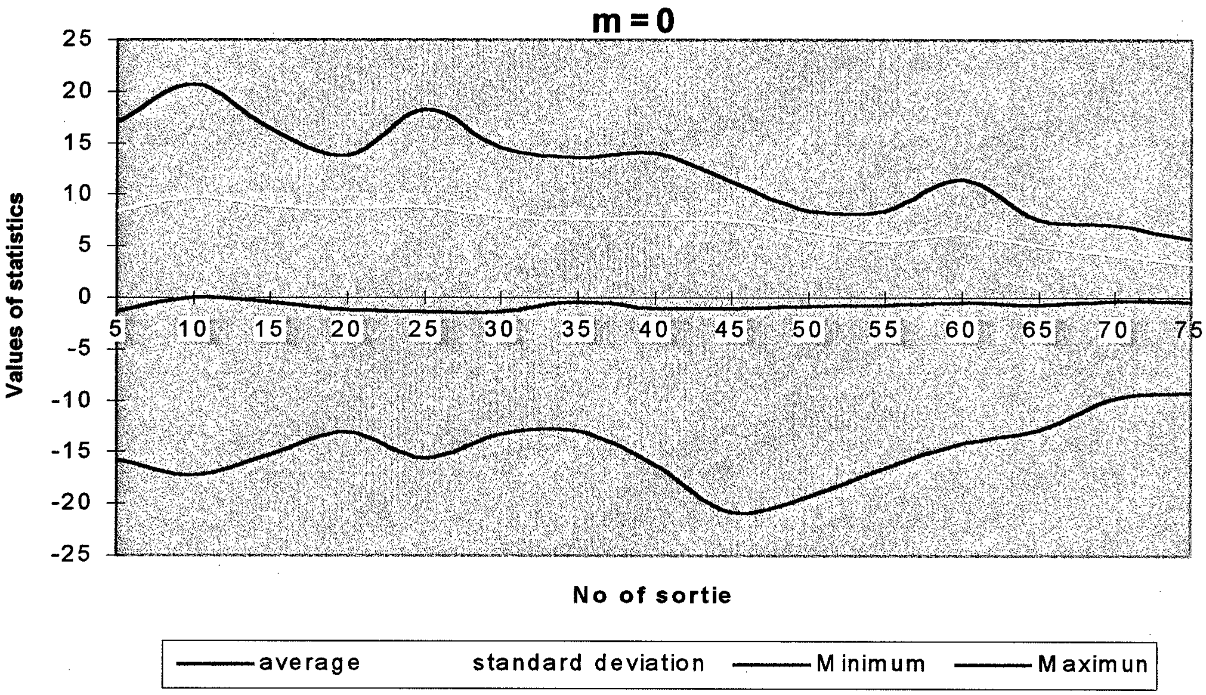


Fig.3.2. Properties of regression model (memory depth *m=0*).

Horizontal axis corresponds to the number of sortie in which the residual performance resource is assessed. *"Average"* is the mathematical expectation of the error of residual performance resource assessment. *"Variance"* is the standard deviation of the error. *"Max"* and *"Min"* values correspond to the maximal and minimal values of the error over the tested trajectories for 25 units.

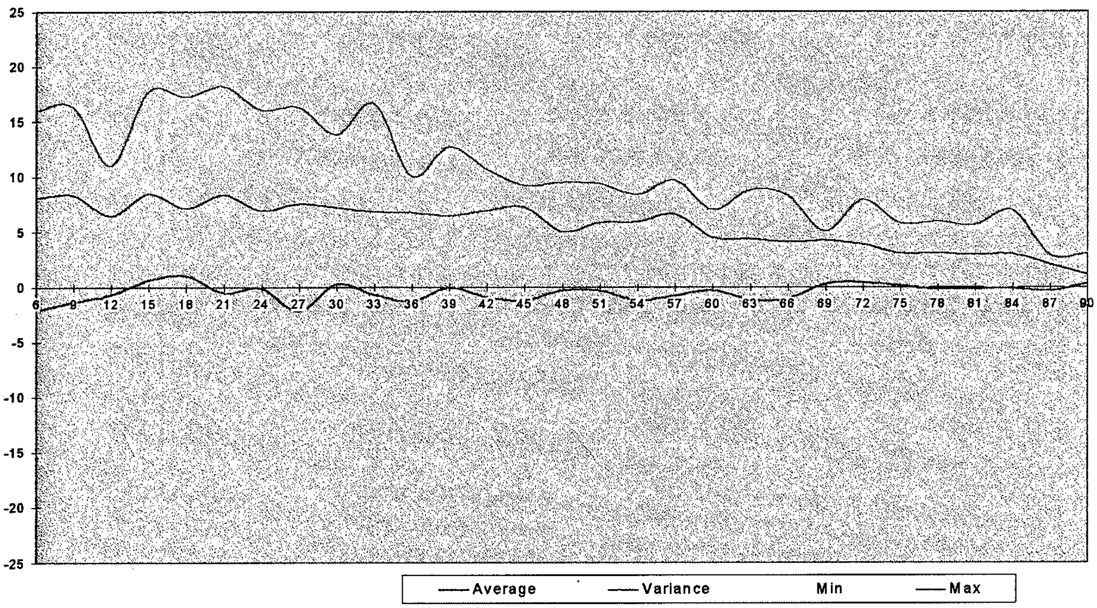


Fig.3.3. Properties of regression model (memory depth *m=1*and *step =3* sorties).

Horizontal axis corresponds to the number of sortie in which the residual performance resource is assessed. *"Average"* is the mathematical expectation of the error of residual performance resource assessment. *"Variance"* is the standard deviation of the error. *"Max"* and *"Min"* values correspond to the maximal and minimal values of the error over the tested trajectories for 25 units.

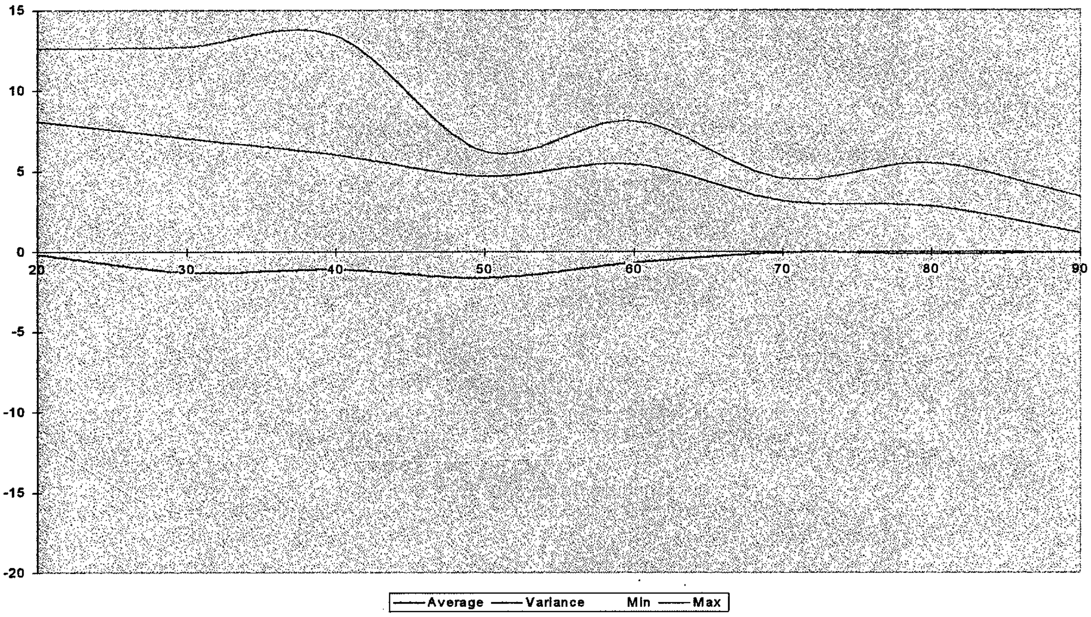


Fig.3.4. Properties of regression model (memory depth *m=1*and *step =10* sorties).

Horizontal axis corresponds to the number of sortie in which the residual performance resource is assessed. *"Average"* is the mathematical expectation of the error of residual performance resource assessment. *"Variance"* is the standard deviation of the error. *"Max"* and *"Min"* values correspond to the maximal and minimal values of the error over the tested trajectories for 25 units.

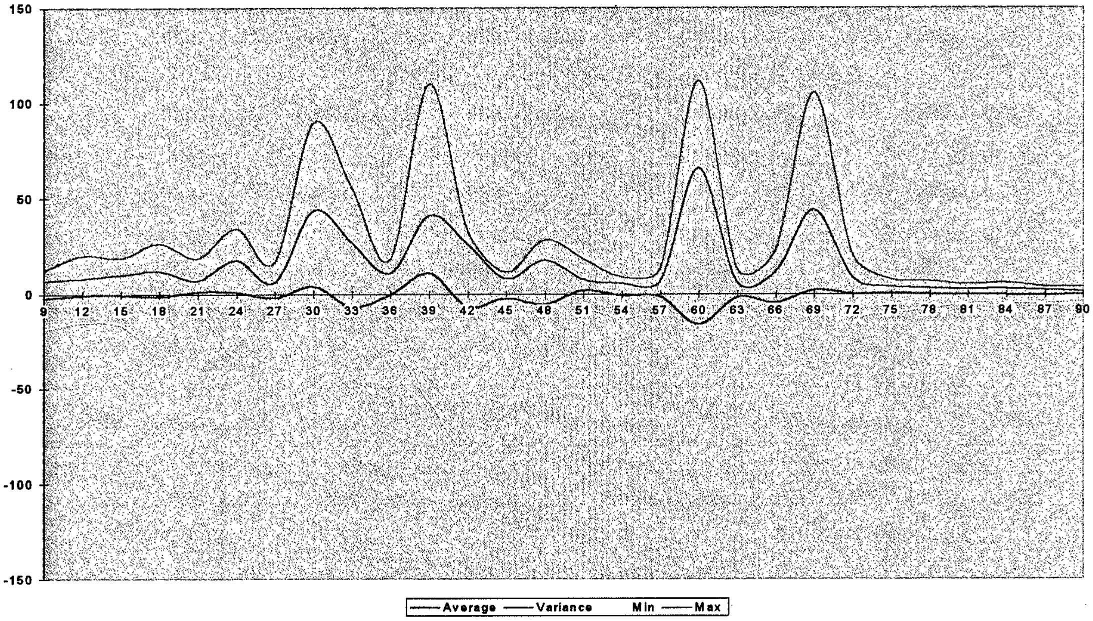


Fig.3.5. Properties of regression model (memory depth *m=2* and *step =3* sorties).

Horizontal axis corresponds to the number of sortie in which the residual performance resource is assessed. *"Average"* is the mathematical expectation of the error of residual performance resource assessment. *"Variance"* is the standard deviation of the error. *"Max"* and *"Min"* values correspond to the maximal and minimal values of the error over the tested trajectories for 25 units.

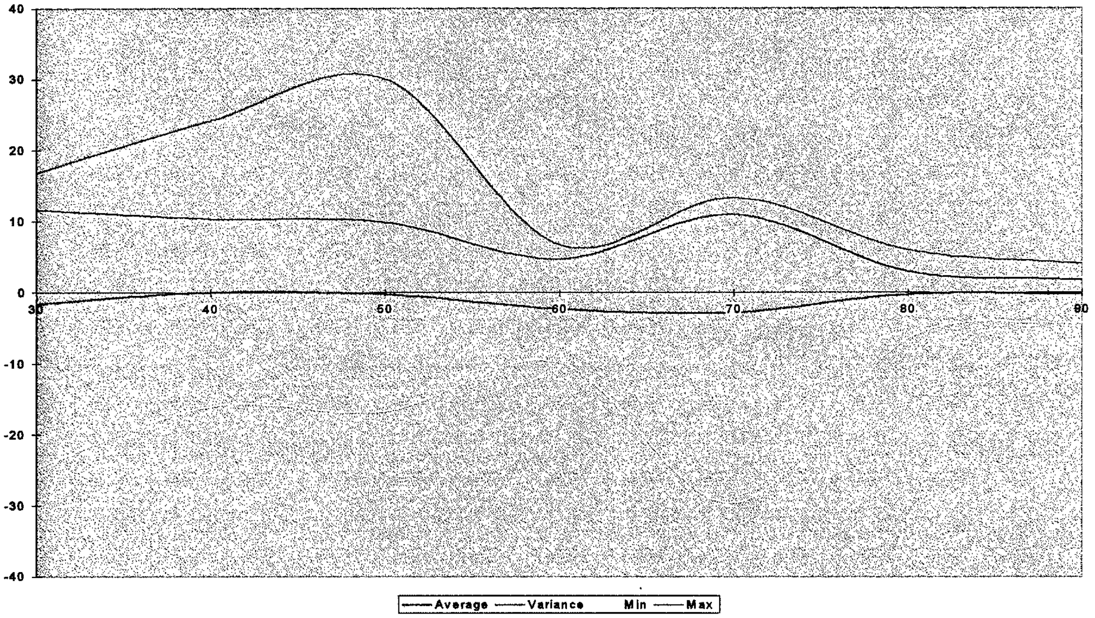


Fig.3.6. Properties of regression model (memory depth *m=2* and *step =10* sorties).

Horizontal axis corresponds to the number of sortie in which the residual performance resource is assessed. *"Average"* is the mathematical expectation of the error of residual performance resource assessment. *"Variance"* is the standard deviation of the error. *"Max"* and *"Min"* values correspond to the maximal and minimal values of the error over the tested trajectories for 25 units.

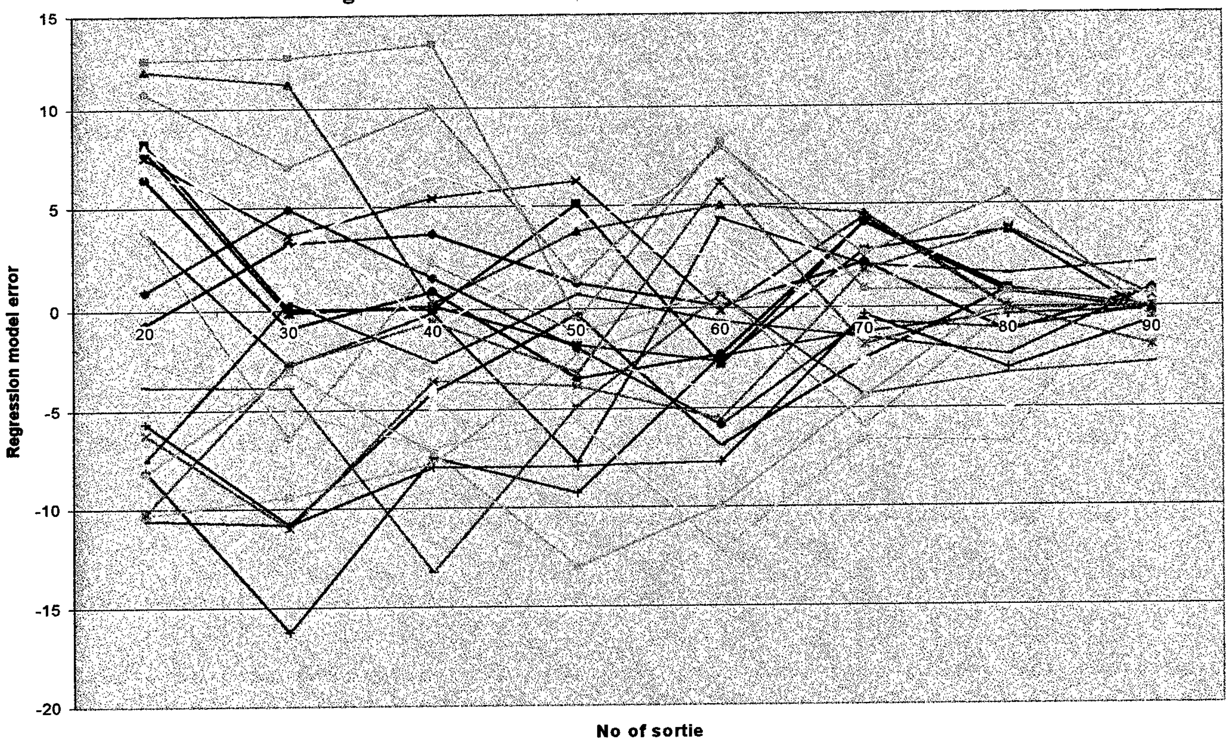


Fig.3.7. Realizations of error of residual performance resource assessment (regression model of memory depth *m=1* and *step =10* sorties).

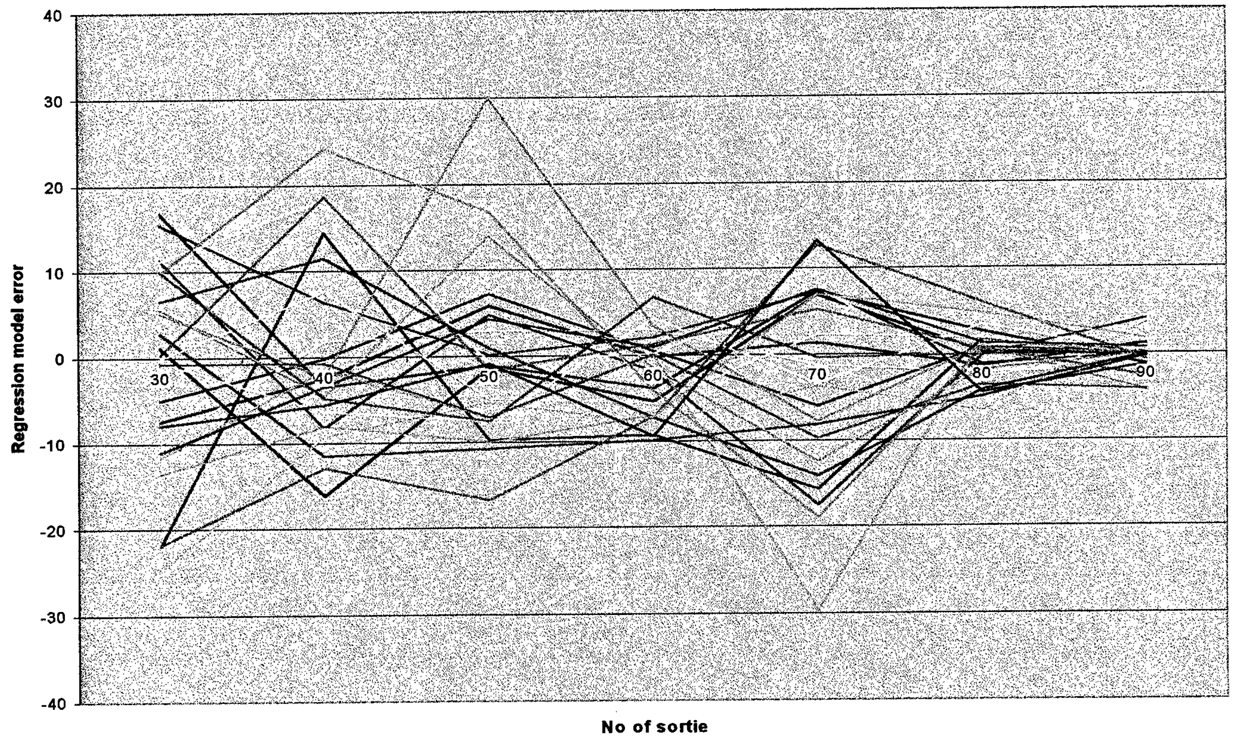


Fig.3.8. Realizations of error of residual performance resource assessment (regression model of memory depth *m=2* and *step =10* sorties).

4. Knowledge Discovery from Statistical Data Base for Health Assessment System Design

4.1. Outline of the Technology of Knowledge Discovery from Statistical Data Base

This section is devoted to the outline of the developed approach that forms a Knowledge Engineering (Knowledge Discovery from Data (KDD)) technology for the prognostic model development. Main ideas of the technology were described in the Interim Report [IR-98]. In this section we in brief repeat the technology description, include some new results obtained in the second and third phases of research and demonstrate the technology numerically on the basis of database generated by DDM described in *Section 2*. It should be noted that this section is self-contained and doesn't require to be familiar to the contents of the Interim Report [IR-98].

In contrast to the data interpretation accepted in the [IR-98], below we consider data divided into three clusters, i.e.

- data records interpreted as "*normal performance" ("no failure"*);
- data records interpreted as *"border-line performance"* (the cases themselves correspond to "*no failure*" but residual performance resource is less than given (chosen) threshold), and
- data records interpreted definitely *as "failure"*.

Let us denote these clusters of data records (statuses of performance) as *"-1"*, *"0"* and *"1"* respectively.

The KDD process aims at the development of a model and a model-based prognostic procedure that provide high quality of classification problem solving and precise prognosis of the probability of failure. The latter is utilized for estimating the probability of failure of a particular module at a given time in the future, say, during a forthcoming sortie, on the basis of the current "history of abuse" of this module.

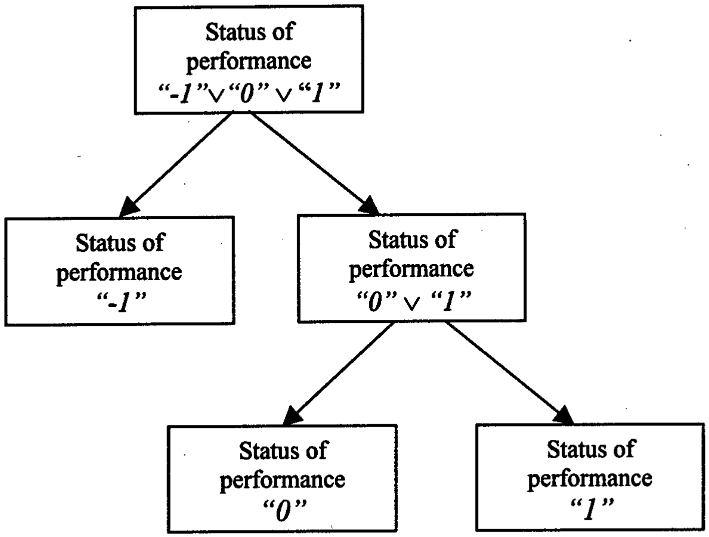
Let us remind that from the Data Mining point of view, this problem constitutes the classification task. A peculiarity of the numerical task considered below compared to one considered in the Interim Report [IR-98] is as follows. In this Report we consider the task that deals with database divided into three clusters as far as in [IR-98] we have considered the binary-status task. This peculiarity necessitates utilization of a multi-step decision making procedure according to the meta–tree depicted below in the fig. 4.1. Let us note that developed software makes it possible to implement search according to this tree for general case when the number of clusters is equal to an arbitrary finite integer value. Each step of search according to the above meta–tree (within a node of meta–tree) corresponds to the technology developed in the Interim Report [IR-98]. Below we describe the upgraded version of this technology.

Let us stress the difference between two notions used in this report, i.e. "decision tree" and "meta-tree". The latter aims at dividing statistical database into two sub-databases via assigning each case the name of meta–class which it belongs to. It is high–level procedure. This procedure doesn't results in any classification predicates. It should be noticed that "meta-tree" design is exclusively the task of expert's responsibility. In contrast, "decision tree" aims at designing classification predicates which forms rules for separation a (meta –) cluster from another one according to "meta–tree. It should be noticed that if we deal with the binary classification task then the corresponding meta–tree consists of the single node.

The Data Mining technology within each node of meta–tree (see, for example, fig.4.1) in the main features corresponds to one developed in [IR-98]. Let us outline it in brief.

Decision tree design includes the development of a partially ordered set of nodes and a predicate dividing a set of cases into two subsets. Consider creating *i-th* node having as input argument a subset of training data S_i which contains cases of two clusters (for example, *'border-line"* and *"failure"*. In general case we deal with two meta-clusters.). The goal of the procedure associated with the node is dividing the set of cases S_i into two clusters in the possibly best way assigning to the node a separation rule and corresponding predicate. This procedure design consists of a number of steps. We consider it below for the node of meta-tree corresponding to the subset of cases "*Status performance*" $\in$ *{"0" ∨ "1"}={"border-line", "failure"}.*

1. *Ranking of two-dimensional subspaces* (*"2-d subspaces"*) of the entire factor space in accordance with the chosen criterion of informativity calculated over training data S_i. It is available for a developer to use a number of heuristic criterions of informativity. *Selection* of the most informative 2-d subspaces is very important task and is solved via user–computer interactions. User-friendly interface makes this task simple for user (see *Subsection* 4.3 below).


Status of performance "-1"∨"0" ∨"1"
Status of performance "-1"
Status of performance "0" ∨ "1"
Status of performance "0"
Status of performance "1"

2. *Visualization* of the projections of both clusters of training data S_i onto the selected 2-d subspace and providing a developer with the opportunity to adjust the separation rule manually using a computer graphical interface. This procedure is a *key point* of the technology. It makes possible to design separation rules of arbitrary shape manually. Note, that the developed software makes it possible to draw manually a *non-linear* and even *non-convex* separation rules and automated generation of the associated predicates.

3. *Division of the experimental data* S_i into two non-overlapping subsets S_i^+, S_i^-, $S_i = S_i^+ \cup S_i^-$. The subset S_i^+ contains the cases of S_i such that the predicate obtained at the previous step is *"true"* over all cases of the set S_i^+, and the subset S_i^- contains the cases of S_i over which it is *"false"*. Based on an additional criterion, each of two subsets S_i^+ and S_i^- is classified as a leaf R_j of the decision tree under development or as its new intermediate node. Decision tree development is ended if it does not contain intermediate nodes that were not processed in accordance with the procedures described above in the steps 1-3.

4. *Each leaf* R_j of the decision tree is mapped to the *predicate* P_j, which is constituted as conjunction of predicates met along the way from root node up to the leaf R_j. In addition, each leaf R_j is mapped to a subset of cases of experimental data for which predicate P_j is *"true"*. Note that each leaf R_j may contain cases of experimental data belonging to both clusters (for example, "*border-line*" and *"failure")* or only to one of them.[1]

Let *r* decision trees are developed. Then the following task is performed to constitute the decision-making procedure.

5. For each decision tree number *k*, *(k=1, 2,..., r), definition of the probabilistic space* in which every leaf R_j^k of decision tree constitutes an elementary event. Each elementary event R_j^k is characterized by a confidence interval probabilities p_k *(X/"border-line")* and p_k *(X/"failure")* estimate defined empirically using training and testing data for any vector of factors $X \in R_j^k$.

[1] As a particular case, we may consider a decision tree that consists only of a root node.

The next two steps correspond to decision making procedure itself and are utilized for model-based estimating (forecasting) the probability of failure of a particular module at a given time on the basis of its current "history of abuse".

6. For given vector of factors *X, definition of the leaf* R_j^k to which vector of factors *X* belongs in each decision tree of number *k,* and hence, definition of the *values of probabilities* p_k *(X/»border-line»)* and p_k *(X/"failure")* (see step 5) obtained by each decision tree. Joint processing of the above probabilities on the basis of so called "Algebraic Bayes' Network" (see [IR-98], [Gorodetski-92], [Gorodetski at al-97]) and calculation of the final values of probabilities *p(X/"border-line")* and *p(X/"failure").*
7. Definition of the *decision making* scheme based on Bayes' approach [Skormin et all-97]. This procedure aims at calculation of the probability of failure *p("failure"/X)* of the device affected by the given cumulative exposures of factors *X.*
8. *Testing* the developed prognostic model and model-based prognostic procedure by using an array of both training and examination data to assess properties of the model and the decision-making procedure.

Of course, the opportunity to utilize our procedure for assessing the probability of failure during the future cycle of the module operation depends upon the ability to forecast exposure of adverse factors at the time in question. The task of development of the appropriate means of forecasting was developed in [Popyack-98].

Note once more that the first five steps result in the definition of a prognostic model. The next two steps constitute a model-based decision-making procedure. The last step is aimed at validation of the resulting model and model-based decision-making procedure.

The mathematical model discussed herein is applicable to solving such practical prognostic related problems as:

- ranking particular environmental conditions as factors responsible for general and particular types of failures,
- determination of particular groups of environmental conditions and assessment of their combined effects on failures in general and on particular types of failures,
- tracking the dynamics of development of cluster models and their statistical characteristics in the process of obtaining new experimental data,
- justification of the development of devices protecting avionics from adverse environmental conditions,
- development of the recommendations on the avoidance of the combined effects of adverse conditions. This could be performed in real-time on the board of an aircraft or spacecraft.

Below we consider the above-described steps of the proposed technology in more details.

4.2. Heuristic Selection of Informative Subspaces. Informativity Criteria

It is well known that learning procedures aimed at extracting knowledge from data are computationally intensive. To decrease the amount of computations, researchers often use heuristic and intuitive notions such as "informativity", "similarity", etc. Formally specified, heuristic and intuitive notions are always problem- or domain-oriented. In our approach we use the intuitive notion of informativity to rank the subspaces of factors (features) and to select a more compressed specification of experimental data for further processing on this basis. We have investigated a number of formal specifications of informativity criteria. All of them can be interpreted as mean square normalized and, possibly, weighted distance between two clusters of statistical data. For large amounts of data the same criteria can be specified in terms of corresponding statistics assessed over data empirically.

Herein and below we use the following notations: $X = \{x_1, x_2, ..., x_n\}$ - is a vector of factors representing cumulative exposure to adverse conditions in hours; *Q∈{"-1", "0", "1"}* - is an integer variable symbolizing the output discrete event (*"normal operation of the device"* corresponds to *Q="-1", Q="0"* corresponds to the *"border-line"* status of the device performance and *"the device failed"* corresponds to *Q="1"*; observed data are indexed by the symbols *r, s;* the number of cases associated with the node *"Status of performance" ={"0"* ∨ *"1"}* is equal to *N*, $N = K_0 + K_1$, where K_0, K_1 are the total number of realizations of cluster *"0"* and cluster *"1"* respectively. Therefore, experimental database consists of the subsets of K_0 realizations of cluster *"0"* marked by superscript *"0"*, for example, $\{x_1^0(r), x_2^0(r), \dots, x_n^0(r)\}$, and K_1 realizations of cluster *"1"* marked by superscript *"1"*, for example, $\{x_1^1(s), x_2^1(s), \dots, x_n^1(s)\}$. Additional notations will be introduced later.

The informativity criteria were selected as the most appropriate due to

- their adequacy to experts' intuitive interpretation of the subspace informativity,
- complexity of the subspace ranking task and
- on the basis of numerical experiments [IR-98].

Criteria (4.1) - (4.3) below correspond to a two-dimensional case but they also can be defined in a subspace of arbitrary dimension[2].

$$\overline{D}^2{}_{l,q} = (K_1K_0)^{-1}\sum_{r=1}^{K_1}\sum_{s=1}^{K_0}\{[x_l^1(r) - x_l^0(s)]^2 / \sigma_l{}^2 + [x_q^1(r) - x_q^0(s)]^2 / \sigma_q{}^2\}, \qquad (4.1)$$

$$\begin{aligned} M[D_{lq}^2] = {} & w_{x_l}(0)/\sigma_l^2 + w_{x_q}(0)/\sigma_q^2 + w_{x_l}(1)/\sigma_l^2 + w_{x_q}(1)/\sigma_q^2 + \\ & + (\Delta\overline{x}_l^{0,1})^2/\sigma_l^2 + (\Delta\overline{x}_q^{0,1})^2/\sigma_q^2 \end{aligned} \qquad (4.2)$$

where σ_l, σ_q - are standard deviations of variables x_l and x_q estimated over the entire range of experimental data, $(\Delta\overline{x}_l^{0,1})^2, (\Delta\overline{x}_q^{0,1})^2$ are squared distances between mathematical expectations of vectors of factors within clusters *"0"* and *"1"* respectively in the subspace comprising factors x_l, x_q.

$$Dw^2{}_{l,q} = (K_1K_0)^{-1}\sum_{r=1}^{K_1}\sum_{s=1}^{K_0} a_s^0 a_r^1\{[x_l^1(r) - x_l^0(s)]^2 / \sigma_l{}^2 + [x_q^1(r) - x_q^0(s)]^2 / \sigma_q{}^2\}, \qquad (4.3)$$

where a_s^0, a_r^1 - are weights assigned to cases (realizations) number *r* and number *s* of clusters *"0"* and *"1"* respectively. Weights a_s^0 and a_r^1 are calculated according to the algorithm given below:

$$\Delta\overline{x}_l = (\overline{x}_l^0 - \overline{x}_l^1)/\sigma_l, \ \Delta\overline{x}_q = (\overline{x}_q^0 - \overline{x}_q^1)/\sigma_q, \ b = \sqrt{(\Delta\overline{x}_l)^2 + (\Delta\overline{x}_q)^2}$$

$$\vec{e}_{lq} = <\Delta\overline{x}_l / b, \ \Delta\overline{x}_q / b> = <\overline{e}_l, \overline{e}_q>$$

For all cases of cluster «1» do (r=1,2,..., K_1)

$$d_r^0 = |\ e_l(x_l^1(r) - \overline{x}_l^0)/\sigma_l + e_q(x_q^1(r) - \overline{x}_q^0)/\sigma_q\ |$$

$$d_r^1 = |\ e_l(x_l^1(r) - \overline{x}_l^1)/\sigma_l + e_q(x_q^1(r) - \overline{x}_q^1)/\sigma_q\ |$$

$$a_r^1 = \begin{cases} 1 \ \mathit{if}\ d_r^0 \geq d_r^1, \\ 0, \ \mathit{if}\ d_r^0 < d_r^1. \end{cases}$$

For all cases of cluster «0» do (s=1,2,..., K_0)

$$d_s^0 = |\ e_l(x_l^0(s) - \overline{x}_l^0)/\sigma_l + e_q(x_q^0(s) - \overline{x}_q^0)/\sigma_q\ |$$

$$d_s^1 = |\ e_l(x_l^0(s) - \overline{x}_l^1)/\sigma_l + e_q(x_q^0(s) - \overline{x}_q^1)/\sigma_q\ |$$

$$a_s^0 = \begin{cases} 1 \ \mathit{if}\ d_s^1 \geq d_s^0, \\ 0, \ \mathit{if}\ d_s^1 < d_s^0 \end{cases}$$

Detailed explanation of the sense of weights a_s^0 and a_r^1 was given in [IR-98].

[2] It was told that due to the use of the notion of meta-tree we reduce the general case of classification task to the case of two clusters. That is why we consider below a binary classification and for numerical demonstration we use cases of clusters *"0"* and *"1"* since separation of these clusters is more difficult task.

Note, that criterion (4.2) is a statistical equivalent of criterion (4.1) and is intended to be used for large amounts of experimental data. Criteria (4.1)–(4.2) are additive and this property makes possible to design an efficient optimization procedure of subspace ranking according to their informativity for any arbitrary dimension. Corresponding algorithm was developed and described in [IR-98]. In the fig.4.2-fig.4.3 the samples of printouts of spaces ordering obtained on the basis of informativity criteria (4.1) and (4.3) respectively are given.

In [IR-98] the computational complexity of the algorithms of calculation of criteria (4.1)–(4.3) was adjusted as well.

4.3. Visual Design of Arbitrary Classification Predicates as a Step Towards a New Technology of Classification Model Design

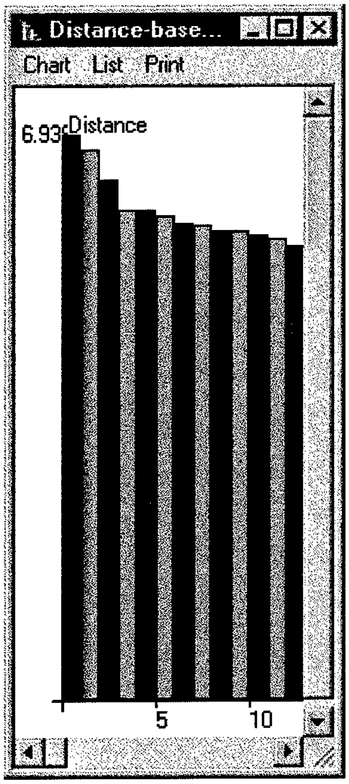
According to the accepted methodology of prognostic model development based on numerical experimental data at the next step so-called classification predicates ([Skormin at al-97], [Skormin at al-99], [IR-98]) are developed. Actually, the meaning of classification predicates introduced below is twofold. First, they form a basis for the definition of prognostic rules. Second, one can consider classification predicates as a feature that represents experimental data on a binary scale instead of the original numeric scale. The latter view is very useful from general Data Mining point of view: since the original experimental data contains both continuous and discrete columns (factors), the utilization of classification predicates facilitates the transformation of all columns of the original experimental data to a discrete format. This transformation is typical in performing Data Mining and KDD tasks but approach considered below is new one and possess a number of very fruitful advantages outlined in the following sections.

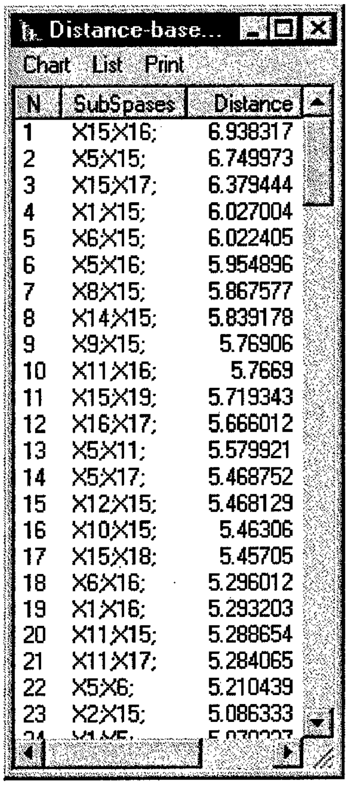
Conceptually, a classification predicate is viewed as the predicate associated with a separation rule designed within a subspace of low dimension. Let us recall that in this study in order to facilitate visualization, we consider only 2-d subspaces.

Let us consider projection of two original clusters of experimental data on a 2-d subspace of factors, i.e. onto a plane as shown in fig. 4.4. Assume that a software tool allows a user to draw linear separation bounds that are perceived as good or optimal. Assume that if a user draws a linear separation bound the software tool automatically generates the linear equation $f(x_l,x_q)$ of the corresponding bound and defines the appropriate predicate as follows:

$$\begin{cases} \text{if } f_k(x_l,x_q)\geq 0 \text{ then } P_k \text{ is } \textit{true} \\ \text{if } f_k(x_l,x_q)<0 \text{ then } P_k \text{ is } \textit{false.} \end{cases} \quad (4.4)$$

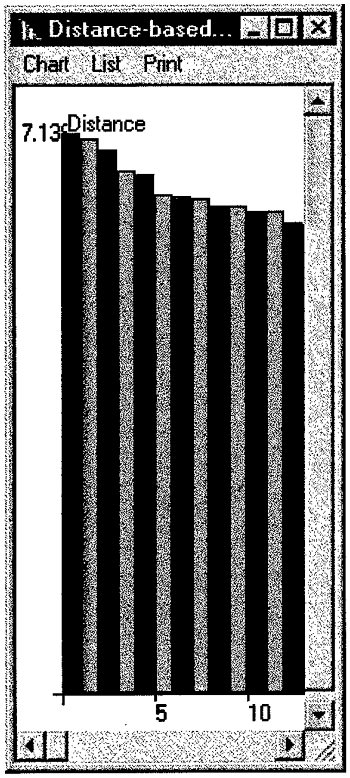
Geometrically, (4.4) implies that in a half-plane $<x_l,x_q>$ predicate P_k is *true* and it is *false* in the alternative half-plane. It is expected that a user has the ability to draw a number of linear separation bounds and corresponding software tool automatically generates equations $f_1(x_l,x_q)$,... ...,$f_m(x_l,x_q)$ and associated predicates P_1, P_2,...,P_m. Each predicate divides plane $<x_l,x_q>$ in two half-planes. Generally, plane $<x_l,x_q>$ will be divided into no more than 2^m convex regions L_i, $i=1,2,...,2^m$ that do not overlap and in combination cover the entire subspace $<x_l,x_q>$. Within each region L_i, $i=1,2,...,2^m$ exactly one conjunction of the length m of predicates P_1, P_2,...,P_m taken with and without negation is *true*. Hence, each region L_i, $i=1,2,...,2^m$ is defined formally by a conjunction of predicates (4.4), an arbitrary combination of such regions is defined by disjunction of above-mentioned conjunctions. One should understand that if a software tool provides a user with the capability to define a number of visually-justified linear separation bounds and associated predicates, then the user is capable to design a very wide class of separation rules. This class contains linear and polygon-like bounds which may correspond to an arbitrary convex and non-convex regions. The latter may be obtained as a combination of convex regions of *truth* of some above mentioned conjunctions. To illustrate this concept let us consider the situation depicted in fig. 4.5.

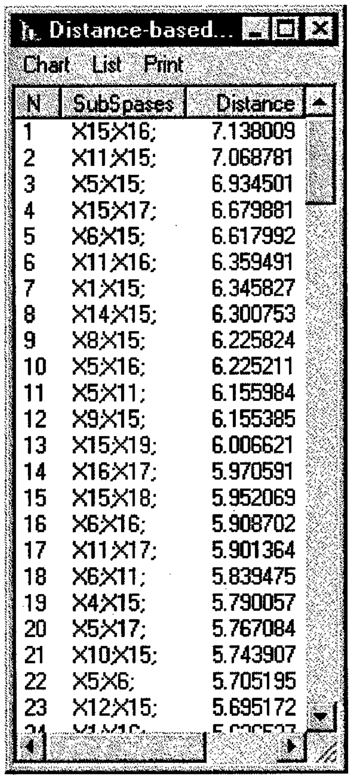




N	SubSpases	Distance
1	X15X16;	6.938317
2	X5X15;	6.749973
3	X15X17;	6.379444
4	X1X15;	6.027004
5	X6X15;	6.022405
6	X5X16;	5.954896
7	X8X15;	5.867577
8	X14X15;	5.839178
9	X9X15;	5.76906
10	X11X16;	5.7669
11	X15X19;	5.719343
12	X16X17;	5.666012
13	X5X11;	5.579921
14	X5X17;	5.468752
15	X12X15;	5.468129
16	X10X15;	5.46306
17	X15X18;	5.45705
18	X6X16;	5.296012
19	X1X16;	5.293203
20	X11X15;	5.288654
21	X11X17;	5.284065
22	X5X6;	5.210439
23	X2X15;	5.086333

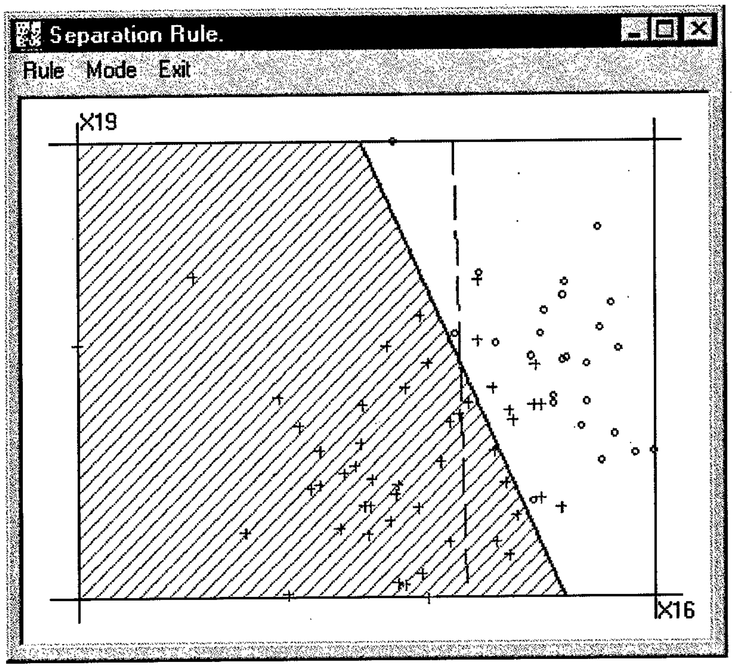
Fig. 4.2. Printouts of histogram and list of 2-dimensional subspaces ordered according to (4.1) measure of informativity.

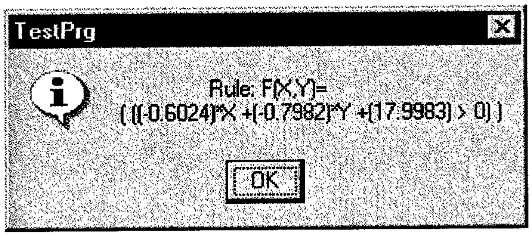




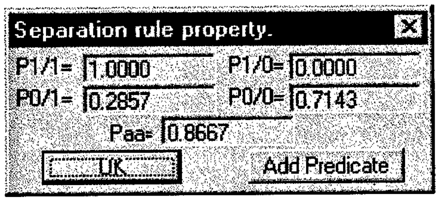
N	SubSpases	Distance
1	X15X16;	7.138009
2	X11X15;	7.068781
3	X5X15;	6.934501
4	X15X17;	6.679881
5	X6X15;	6.617992
6	X11X16;	6.359491
7	X1X15;	6.345827
8	X14X15;	6.300753
9	X8X15;	6.225824
10	X5X16;	6.225211
11	X5X11;	6.155984
12	X9X15;	6.155385
13	X15X19;	6.006621
14	X16X17;	5.970591
15	X15X18;	5.952069
16	X6X16;	5.908702
17	X11X17;	5.901364
18	X6X11;	5.839475
19	X4X15;	5.790057
20	X5X17;	5.767084
21	X10X15;	5.743907
22	X5X6;	5.705195
23	X12X15;	5.695172

Fig. 4.3. Printouts of histogram and list of 2-dimensional subspaces ordered according to (4.3) measure of informativity.



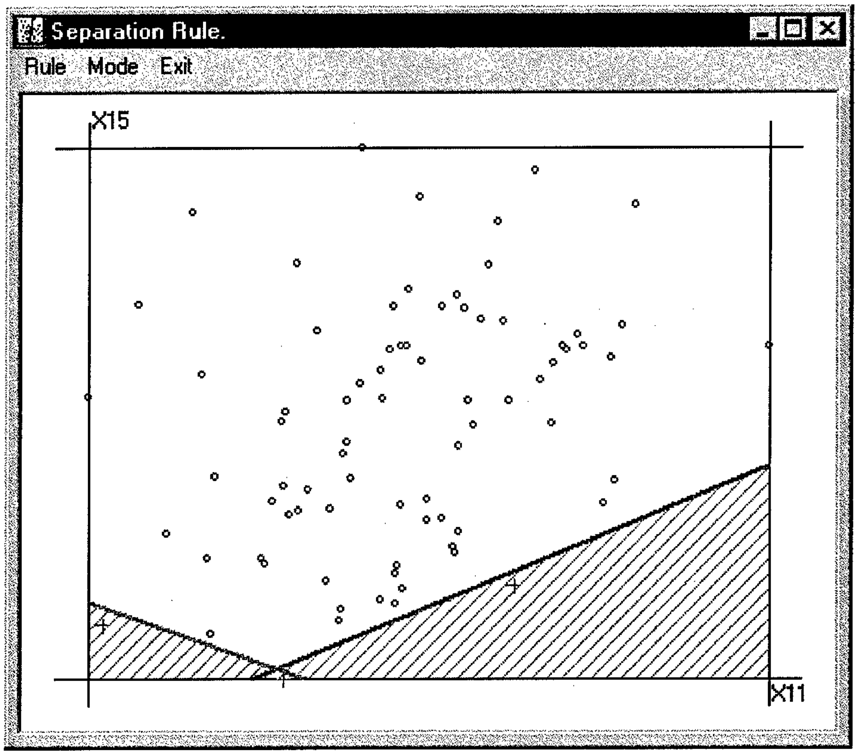


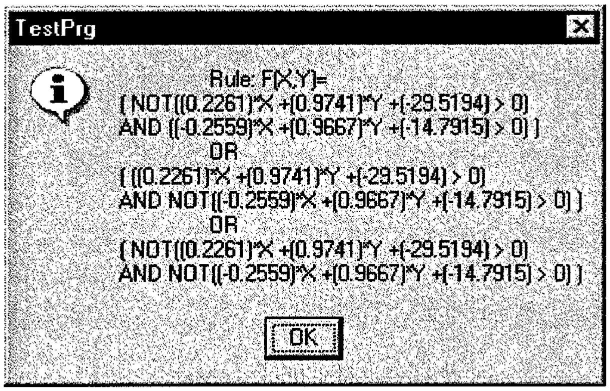
Rule: F(X,Y)=
(((-0.6024)*X +(-0.7982)*Y +(17.9983) > 0))



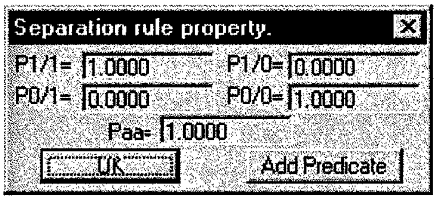
P1/1= 1.0000 P1/0= 0.0000
P0/1= 0.2857 P0/0= 0.7143
Paa= 0.8667

Fig. 4.4. Printouts of (1) visual interface for drawing linear separation boundary (top picture), (2) automatically generated classification predicate (middle picture) and (3) results of assessment of quality (probabilistic properties) of the generated classification predicate (bottom picture). Cases of different clusters are denoted by signs of different colors.





```
        Rule: F(X,Y)=
( NOT((0.2261)*X +(0.9741)*Y +(-29.5194) > 0)
AND ((-0.2559)*X +(0.9667)*Y +(-14.7915) > 0) )
        OR
( ((0.2261)*X +(0.9741)*Y +(-29.5194) > 0)
AND NOT((-0.2559)*X +(0.9667)*Y +(-14.7915) > 0) )
        OR
( NOT((0.2261)*X +(0.9741)*Y +(-29.5194) > 0)
AND NOT((-0.2559)*X +(0.9667)*Y +(-14.7915) > 0) )
```



P1/1= 1.0000 P1/0= 0.0000
P0/1= 0.0000 P0/0= 1.0000
Paa= 1.0000

Fig. 4.5. Printouts of (1) visual interface for drawing polygon-like separation boundary (top picture), (2) automatically generated non-linear classification predicate (middle picture) and (3) results of assessment of quality (probabilistic properties) of the generated classification predicate (bottom picture). Cases of different clusters are denoted by signs of different colors.

Assume that a user defines three linear separation bounds by visualizing the computer-generated clustering pattern in the plane $<x_{11}, x_{15}>$. The predicates P_1 and P_2 associated with these bounds are represented formally as follows:

$$P_1 = (0.226X_{11} + 0.947X_{15} - 29.52 \geq 0),$$
$$P_2 = (-0.256X_{11} + 0.967X_{15} - 14.79 \geq 0).$$

Both of them are assigned by value of *"true"* within the half-plane located above from the corresponding linear bounds. Hence, the white region corresponds to the *truth* domain of both predicates P_1 and P_2. The green colored region is non-linear and non-convex and is specified by the following logic formula given over predicates P_1 and P_2:

$$CP_1 = (\neg P_1 \& P_2) \vee (P_1 \& \neg P_2) \vee (\neg P_1 \& \neg P_2)$$

The separation bound and the corresponding predicate CP is designed usually to separate cases of two clusters as good as possible. The last requirement can be specified formally as follows:

$$N_{11} > M_{10},\ M_{00} > N_{01}, \tag{4.5}$$

where N_{11} - is the number of realizations of cluster *"1"* for which predicate CP_1 is assigned the value *"true"; N_{01}* is the number of realizations of cluster *"1"* for which predicate CP_1 is assigned the value *"false"*; M_{10} is the number of realizations of cluster *"0"* for which predicate CP_1 is assigned the value *"true"*, and M_{00} is the number of realizations of cluster *"0"* for which predicate CP_1 is assigned the value *"false"*. It is clear that $(N_{11} + M_{00})$ realizations of experimental data are correctly classified by predicate CP, and $(M_{10} + N_{01})$ realizations of data are classified by predicate CP_1 erroneously. For example, these numbers for predicate CP_1 are (see fig.4.5):

$$N_{11}=77,\ M_{10}=0,\ N_{01}=0,\ M_{00}=3.$$

Definition 1. Predicate that meets condition (4.4) and inequalities (4.5) is a classification predicate.

Definition 1 is non-formal and introduces the term that is used elsewhere.

Based on experimental data every classification predicate CP_k, *k=1,2, ..., m,* can be assigned a number of attributes that represent the quality of classification that it is expected to achieve. Let us consider empirical estimates of probabilities of the correct and erroneous classifications of realizations of experimental data represented as a matrix:

$$p(CP_k) = \begin{bmatrix} p_k(1/1) & p_k(1/0) \\ p_k(0/1) & p_k(0/0) \end{bmatrix}, \tag{4.6}$$

Note that the first argument within the brackets corresponds to the decision made by classification predicate, and the second argument corresponds to the actual status of the realization in question. These estimates can be calculated as follows:

$$p_k(1/1)=N_{11}(k)/[N_{11}(k)+M_{10}(k)], \quad p_k(1/0)=M_{10}(k)/[N_{11}(k)+M_{10}(k)],$$
$$p_k(0/1)=N_{01}(k)/[M_{00}(k)+N_{01}(k)], \quad p_k(0/0)=M_{00}(k)/[N_{01}(k)+M_{00}(k)],$$

For example, for classification predicate CP_1 (see fig.4.2) the above estimates are as follows:

$$p_k(1/1)=1;\ p_k(1/0)=0$$
$$p_k(0/1)=0;\ p_k(0/0)=1.$$

In fig. 4.6, fig.4.8, fig.4.9 and fig.4.10 one can see a number of samples of graphical synthesis of separation rules and corresponding classification predicates embedded in decision trees design (see *Subsections 4.4* and *4.5*). They were designed for the case study of the developed technology considered below. Compared to the Interim Report [IR-98] a peculiarity of case study is that it is based on statistical database generated by DDM developed in *Section 2*.

Let us note that one more approach to the classification predicates synthesis was developed by V.Skormin [Skormin at al-99]. In contrast to the above polygon-like separation rule design that is natively interactive he developed an "automatic procedure resulting in ellipse-based separation rules. This approach is efficient enough but comparing to the former approach it has to be implemented in the "same style", i.e. as a procedure supported by interactive visualizing software. This comparison is a goal of the future work in the framework of Data Mining oriented research.

4.4. Forest of Decision Trees as a Step Towards Improving Quality of Classification

The general idea of decision tree was outlined in *Section 4.1.* The decision tree development procedure aims at finding the non-correlated informative subspaces over the training subset of experimental data. Let us explain the term "correlation" within the context of the Report.

Let CP_1 and CP_2 be the classification predicates associated with a selected pair of the most informative subspaces. Let each of them divide the entire set of cases S into two subsets, i.e. into subspaces $\{S_1^1, S_2^1\}$ and into subspaces $\{S_1^2, S_2^2\}$ respectively. It may turn out that $S_1^1 \approx S_1^2$ and $S_1^2 \approx S_2^2$. If such non-formal equalities are held then the second informative subspace is not able to improve remarkably the classification procedure. Hence, the second subspace is informative in itself but is not informative if it is added to the first one. This example explains approximately in which sense we use the term "subspaces correlation". The latter may be specified formally but it is not necessary because it is clear how to take into account correlation of subspaces in the interactive visualized selection procedure to avoid utilization of highly "correlated" subspaces in the classification procedure. Let us explain how it can be performed.

In the Fig.4.6, a sample of a decision tree is depicted. In the first step the developer has selected the 2-d subspace $<x_5, x_{12}>$ as the most informative. The upper screen (Fig.4.6) depicts projections of both clusters *"0"* and *"1"* onto the plane $<x_5, x_{12}>$ and a linear separation rule (continuous line) established by an expert manually as the optimal one (The broken line corresponds to the separation rule calculated automatically but rejected by the developer.). This separation rule divides the entire set of training experimental data into two non-overlapping subsets S_1, S_2.

The next steps of subspace selection are applied separately to subsets S_1 and S_2 which form two new nodes of decision tree. Therefore, on the second and third steps we have to solve two tasks of informative subspaces selection for the two above mentioned subsets S_1 and S_2 of experimental data. In the Fig.4.6 the results of the second step selections are visualized and represented as printouts. The next and all further steps can be realized in the same way. As a result, a decision tree depicted in the Fig.4.6 is obtained. This decision tree is provided by all related information, i.e. set of 2-d subspaces, equations of separation rules, classification predicates, related probabilities, etc., obtained by the developed software automatically (This information is omitted in the fig.4.6). One can see that some separation bounds are chosen in a nonlinear form.

4.5. Probabilistic decision making procedure

The result of the above procedure is the decision tree such that each its leaf is mapped to a subset of cases of original training data. These subsets are not overlapping and their union covers the entire set of training data. On the other hand, each leaf is mapped to its own classification predicate constituted as conjunction of classification predicates of decision tree nodes met along the way from tree root up to respective leaf. Each predicate is true in the concrete region of the factor space, and the regions corresponding to the different leaves are not overlapping and cover the entire factor space. Hence, they can be used as the elementary events to design a probabilistic space. Let us consider in more formal way how the probabilistic space is constituted and how it is utilized to assess the probability of failure of a device having given "history of abuse".

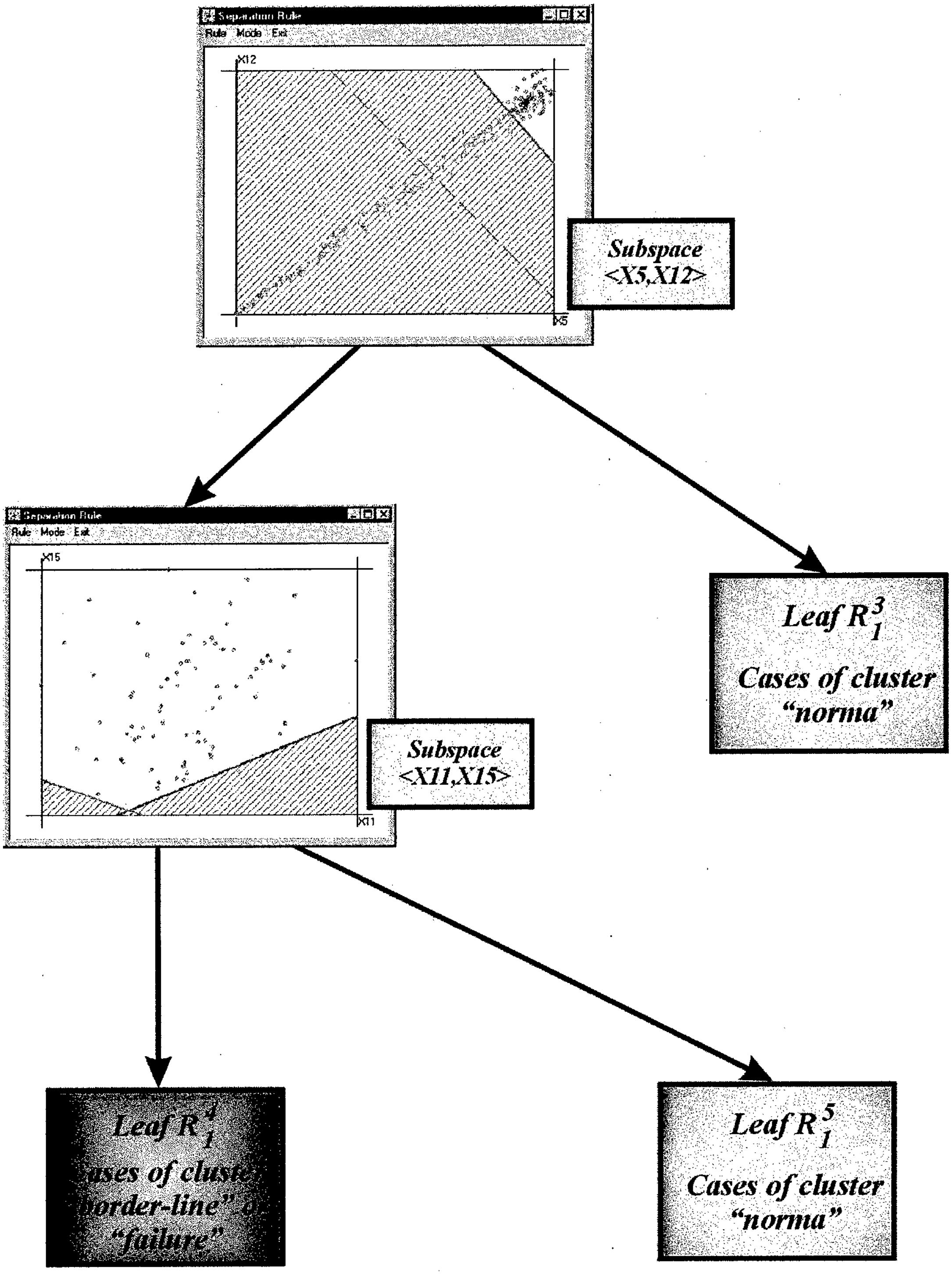


Fig. 4..6. Decision tree for the node of meta-tree <*Status of performance "–1"*∨*"0"*∨*"1"*> (see fig. 4.1)

Let $\{R_1, R_2, ..., R_S\}$ is the set of elementary events which are mapped to the set of leaves of a designed decision tree and $\{P_1, P_2, ..., P_S\}$ is the set of the respective (mapped to corresponding leaves of decision tree) classification predicates. Each such elementary event $R_i \in \{R_1, R_2, ..., R_S\}$ can be mapped to an empirically estimated probability $p_{R_i}(X) = p_i(X)$ on the basis of testing of designed decision tree over both training and testing data in a traditional way. We suppose that these estimations are calculated as confidence intervals for given level of confidence probability. In the same way, confidence intervals of probabilities $p_i(X/0)$ and $p_i(X/1)$ can be estimated. Note that $p_i(X) = p_i(X/0) + p_i(X/1)$. As a result for each predicate P_1, P_2, ..., P_S the following probabilities will be estimated by confidence intervals:

$$p_h(1/1) = p[P_h(X) = "true"/X \in "1"],\ p_h(1/0) = p[P_h(X) = "true"/X \in "0"], \qquad (4.7)$$
$$p_h(0/1) = p[P_h(X) = "false"/X \in "1"],\ p_h(0/0) = p[P_h(X) = "false"/X \in "0"].$$

The availability of the probabilities (4.7) makes it possible to calculate the target probability of failure for any point of the factor space X^* subject to the condition that $P_h(X^*) = "true"$ using Bayes' formula as follows:

$$p[1/P_h(X^*)] = \frac{p(1)p_h(1/1)}{p(1)p_h(1/1) + p(0)p_h(1/0)}. \qquad (4.8)$$

Here $p(1)$, $p(0)$ - are prior probabilities of *"border-line"* and *"failure"* of the device operation. It is obvious that probability of normal operation under the adverse exposures X^* is

$$p_h(0/P_h(X^*)] = 1 - p[1/P_h(X^*)]$$

In the general case when the available size of training and testing data is too small we are not able to obtain the satisfactory accuracy of estimations of task related probabilities to forecast a probability of failure. However, the quality of designed prognostic model depends critically on the above accuracy. Therefore, we have to undertake special efforts to provide the needed accuracy. In the next section we investigate an approach to cope with the above problem.

4.6 Improvement of Assessment of the Task-related Probabilities

Recall that the main goal of the prognostic model under development is the reevaluation of the probability of failure of an avionics module on the basis of its actual "history of abuse" represented by the vector of adverse exposures. It is elsewhere adopted that the accuracy of the assessment of this probability depends on two factors:

(1) the total amount of experimental data that is used for training (prognostic model design) and testing (evaluation of the model quality),

(2) quality of prognostic model and model-based decision-making procedure.

We assume that in our case the amount of experimental data is small[3] and we have no information about the distribution of realizations within data clusters. Therefore, in a general case even the best prognostic model may such that it is not able to provide precise probability assessment. However, we are able to assign a confidence interval for the probability of each elementary event associated with each leaf of the designed decision tree that constitute a guaranteed estimation.

To narrow these confidence intervals and, hence, to improve the accuracy of probability assessment, we proposed the following two approaches:

- Use of a set of prognostic models (a set of decision trees) that differ in factor subspaces utilized by each decision tree of the respective prognostic model. This approach leads to a collective of decision-making procedures. This way is based on the redundancy of information involved in a decision making and results in the improvement of the probability accuracy estimation. To realize such approach it is necessary to develop a special algorithm for joint processing of probabilities resulting

[3] Otherwise the problem of improving of the probabilities assessment doesn't exists.

from each decision tree. We proposed to use the algorithm based on the Algebraic Bayes' Network (ABN) approach developed in [Gorodetski-92], [Gorodetski et al-97].

- It is proposed to replace the traditional Bayes' formula (13) resulting in a posterior probability by its equivalent developed on the basis of methods of interval mathematics which facilitates the calculation of a guaranteed estimate of the probability of failure.

Consider the brief discussion of the algorithms implementing the above approaches. In [IR-98] these algorithms were illustrated by a numerical example of the assessment of failure probability of an avionics module. It is understood that the quantities generated by a decision-making procedure represented by a decision tree depend on the choice of the root and intermediate nodes. Typically, we are able to design a number of decision trees utilizing different subspaces of the factor space that results in information redundancy. Consider the impact of this redundancy on the accuracy of the prognostic procedure.

Assume that a set of three decision trees has been established within a particular prognostic model. Consider the application of this model for the assessment of the probability of failure of an avionics module subjected to adverse exposures X. Assume that vector X results in elementary events R_i^1, R_j^2 and R_k^3 or, using some specific jargon, belongs to the appropriate leaves of the first, second and third decision trees. The respective probabilities are defined as

$$R_i^1:\ a_i^1 \le p(X) \le b_i^1,\ a_i^1(0) \le p(X/0) \le b_i^1(0),\ a_i^1(1) \le p(X/1) \le b_i^1(1); \tag{4.9}$$

$$R_j^2:\ a_i^2 \le p(X) \le b_j^2,\ a_j^2(0) \le p(X/0) \le b_j^2(0),\ a_j^2(1) \le p(X/1) \le b_j^2(1); \tag{4.10}$$

$$R_k^3:\ a_k^3 \le p(X) \le b_k^3,\ a_k^3(0) \le p(X/0) \le b_k^3(0),\ a_k^3(1) \le p(X/1) \le b_k^3(1). \tag{4.11}$$

It could be seen that vector X belongs to all tree subspaces, therefore the probability of the event $R_i^1 \wedge R_j^2 \wedge R_k^3$, i.e. the probability of the event $p(R_i^1 \wedge R_j^2 \wedge R_k^3)$ could be defined on the basis of Algebraic Bayes' Network (ABN) approach developed in [Gorodetski et al-97] (see also *Section 5*). This approach reflects the basics of the probability theory and requires that the interval constraints (4.9) - (4.11) be supplemented by some fundamental axioms.

In the case under consideration, we specify the interrelationships between probabilities that are defined by each decision tree (they are given in (4.9) - (4.11)) and the probability $p(R_i^1 \wedge R_j^2 \wedge R_k^3)$. Following the ABN approach we represent the interrelationships between probabilities by a Hasse diagram [Birkhoff-67] as shown in fig.4.7. Denote the probabilities of events constituted by intersections of events R_i^1, R_j^2 and R_k^3 as follows:

$$p(X\in R_i^1)=p(X_1),\ p(X\in R_j^2)=p(X_2),\ p(X\in R_k^3)=p(X_3),$$

$$p[(X\in R_i^1)\&(X\in R_j^2)]=p(X_1X_2),\ p[(X\in R_i^1)\&(X\in R_k^3)]=p(X_1X_3),$$

$$p[(X\in R_j^2)\&(X\in R_k^3)]=p(X_2X_3),\ p[(X\in R_i^1)\&(X\in R_j^2)\&(X\in R_k^3)]=p(X_1X_2X_3).$$

The following are the interrelationships between probabilities reflecting the axioms of norm and additivity:

$$\left.\begin{array}{ll}
p(X_1)+p(X_2)-\ p(X_1X_2) \le 1 & p(X_1)+p(X_2)+p(X_3)-p(X_1X_2)-p(X_1X_3)- \\
p(X_1)+p(X_3)-p(X_1X_3) \le 1 & -p(X_2X_3)+p(X_1X_2X_3) \le 1. \\
p(X_2)+p(X_3)-p(X_2X_3) \le 1, & p(X_2X_3)-p(X_1X_2X_3) \ge 0, \\
p(X_2)-p(X_1X_2) \ge 0 & p(X_1X_3)-p(X_1X_2X_3) \ge 0, \\
p(X_1)-p(X_1X_2) \ge 0 & p(X_1X_2)-p(X_1X_2X_3) \ge 0, \\
p(X_3)-p(X_1X_3) \ge 0 & \\
p(X_1)-p(X_1X_3) \ge 0 & p(X_3)-p(X_1X_3)-p(X_2X_3)+p(X_1X_2X_3) \ge 0, \\
p(X_3)-p(X_2X_3) \ge 0 & p(X_2)-p(X_1X_2)-p(X_2X_3)+p(X_1X_2X_3) \ge 0, \\
p(X_2)-p(X_2X_3) \ge 0 & p(X_1)-p(X_1X_2)-p(X_1X_3)+p(X_1X_2X_3) \ge 0,
\end{array}\right\} \tag{4.12}$$

and, as usual, $p(X_i) \geq 0,\ i = 1,2,3,\ p(X_i X_j) \geq 0,\ i,j = 1,2,3,\ \ p(X_1 X_2 X_3) \geq 0$.

These interrelationships are viewed as the *background knowledge*, that could be incorporated in the estimation of probability $p(X) = p(X_1 X_2 X_3)$ in the form of two linear programming problems as follows:

1. min{ $p(X_1 X_2 X_3)$*)* *under constraints (4.9), (4.10), (4.11) and (4.12) (lower bound);* (4.13)

2. max{ $p(X_1 X_2 X_3)$*)* *under constraints (4.9), (4.10), (4.11) and (4.12); (upper bound).* (4.14)

Note that based on experimental data it is possible to assess intervals for probabilities $p(X_1 X_2)$, $p(X_1 X_3)$ and $p(X_2 X_3)$ as well. This can be useful for further narrowing intervals of probabilities in question. This approach to improving the accuracy of interval probabilities is very fruitful. It was demonstrated numerically by example in [IR-98].

Now consider the use of Bayes' formula (4.8) for calculation of the posterior probability *p(1/X)* in the case when probabilities *p(X/1)* and *p(X/0)* are given by their confidence intervals. Our goal is to calculate the upper bound of probability *p(1/X)*. Therefore, the task is to find its maximum value subject to constraints

$$a_1 \leq P(X / 1) \leq b_1,\ a_0 \leq P(X / 0) \leq b_0. \qquad (4.15)$$

Simple analysis of this optimization task shows that it is equivalent to the task of maximization of the quotient *p(X/1)/p(X/0)*. This provides the justification for the following formula:

$$max\{p(1 / X) = \frac{p(1)b_1}{p(1)b_1 + p(0)a_0} \qquad (4.16)$$

The detailed numerical demonstration of the developed technology for design information-based health assessment system was given in [IR-98]. Additional numerical results obtained for statistical database generated by the developed DDM is given in the next subsection. Statistical database itself contained learning and testing Data is given in *Appendix A2.*

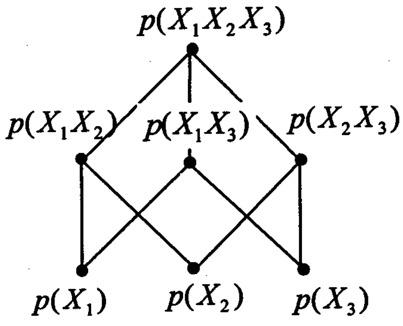


Fig.4.7. Algebraic Bayes' Network for three- propositional case

4.7. Numerical results

In [IR-98] we demonstrated the developed technology of probability of failure assessment and prognosis in detail. In this report we omitted this. Instead, let us consider additional numerical results regarding the statistical assessment of the quality of decision making procedure over the testing database. The latter was generated on the basis of DDM model developed in this research and described in *Section 1.* This database consists of 200 cases including *125* cases of the cluster *"no failure"*, *50* cases of the cluster *"border-line"* and *25* cases of the cluster *"failure"* (see *Appendix A2*). Note that these cases were not involved in the learning procedures resulting in decision trees depicted in the fig.4.6, fig.4.8, fig.4.9 and fig.4.10.

Numerical results of testing are presented in tab.4.1 below. One can see that each decision tree provides the sufficient level of classification quality. Note that these results were obtained on the basis of small training database. In fact, if we were continuing learning procedure involving in this process new training data we should be able to reach much more high level of perfect classification.

Let us comment data in two last columns of the table. The column #7 contains probabilities of classification (the latter are presented in the column #2) for the case of using decision procedure on the basis of decision trees voting according to the rule "two of three". In the column #8 we presented probabilities of classification for the case if voting is organized corresponding to the rule "if at least one decision tree votes for "failure" then decision is "failure". Note that in this Report we don't demonstrate numerically use of ABN because the latter was presented in Interim report [IR-98].

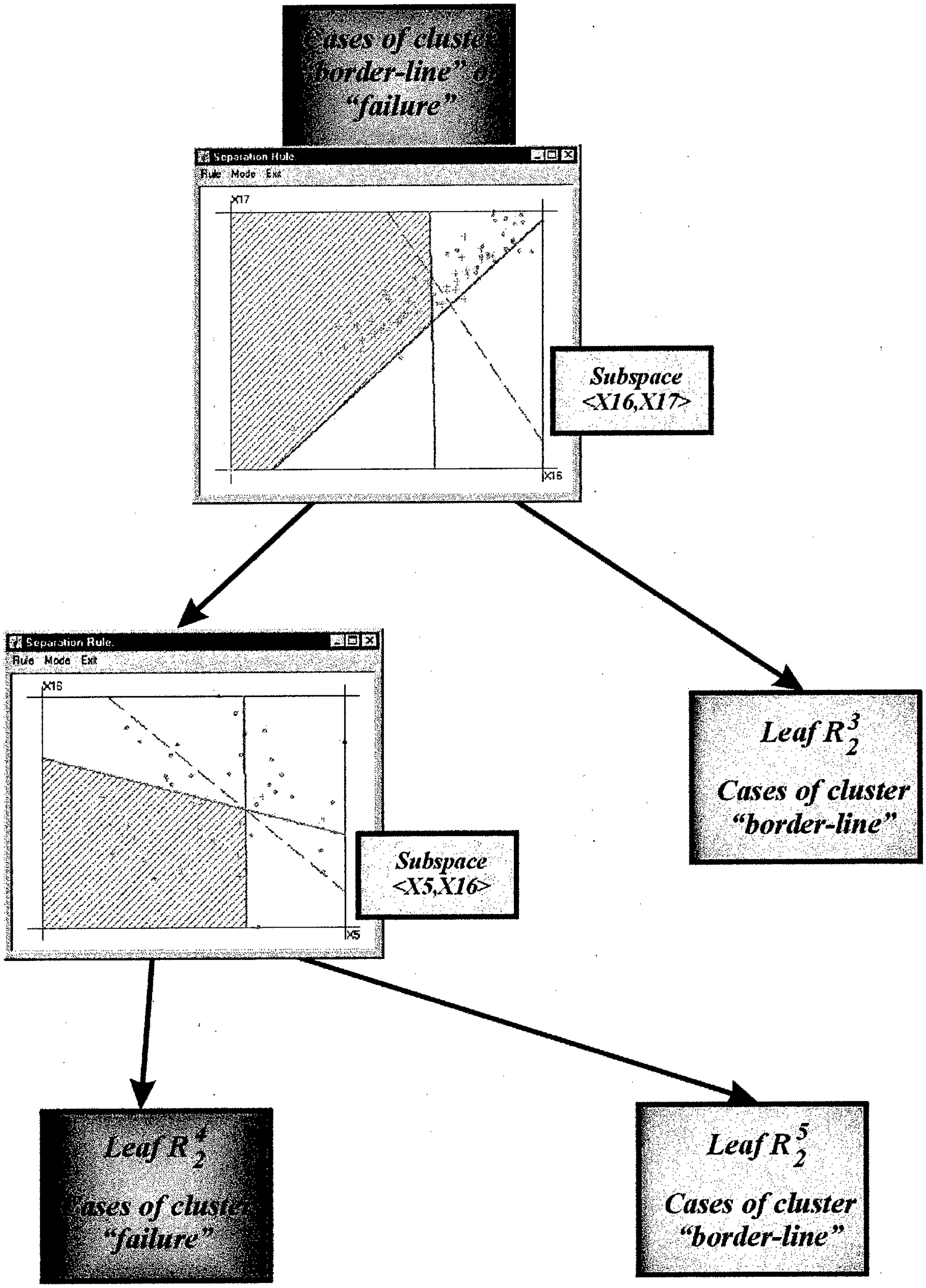


Fig. 4.8. Decision tree #1 for the node of meta-tree <*Status of performance "0" ∨ "1"*> (see fig. 4.1)

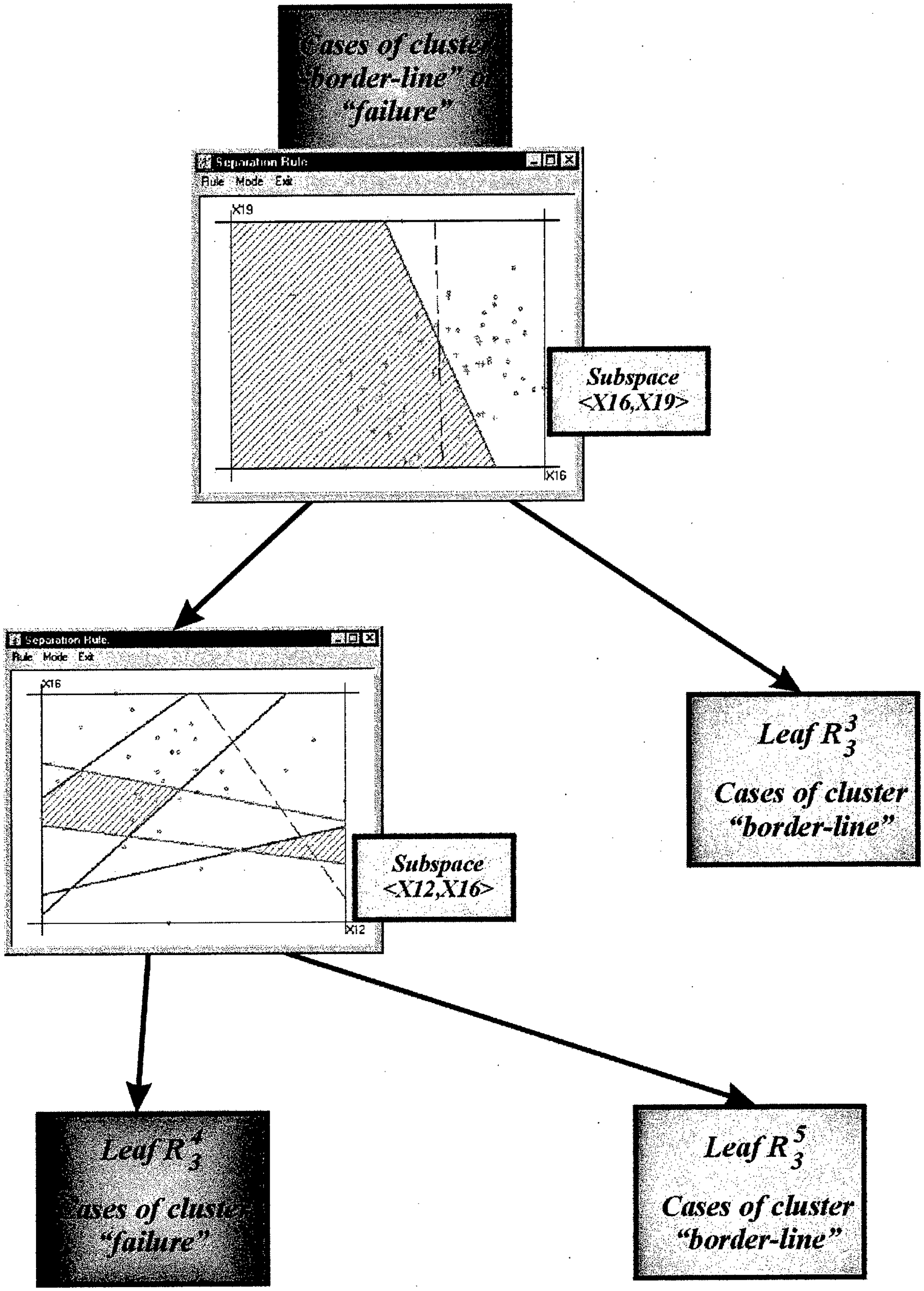


Fig. 4.9. Decision tree #2 for the node of meta-tree <*Status of performance "0" ∨ "1"*> (see fig. 4.1)

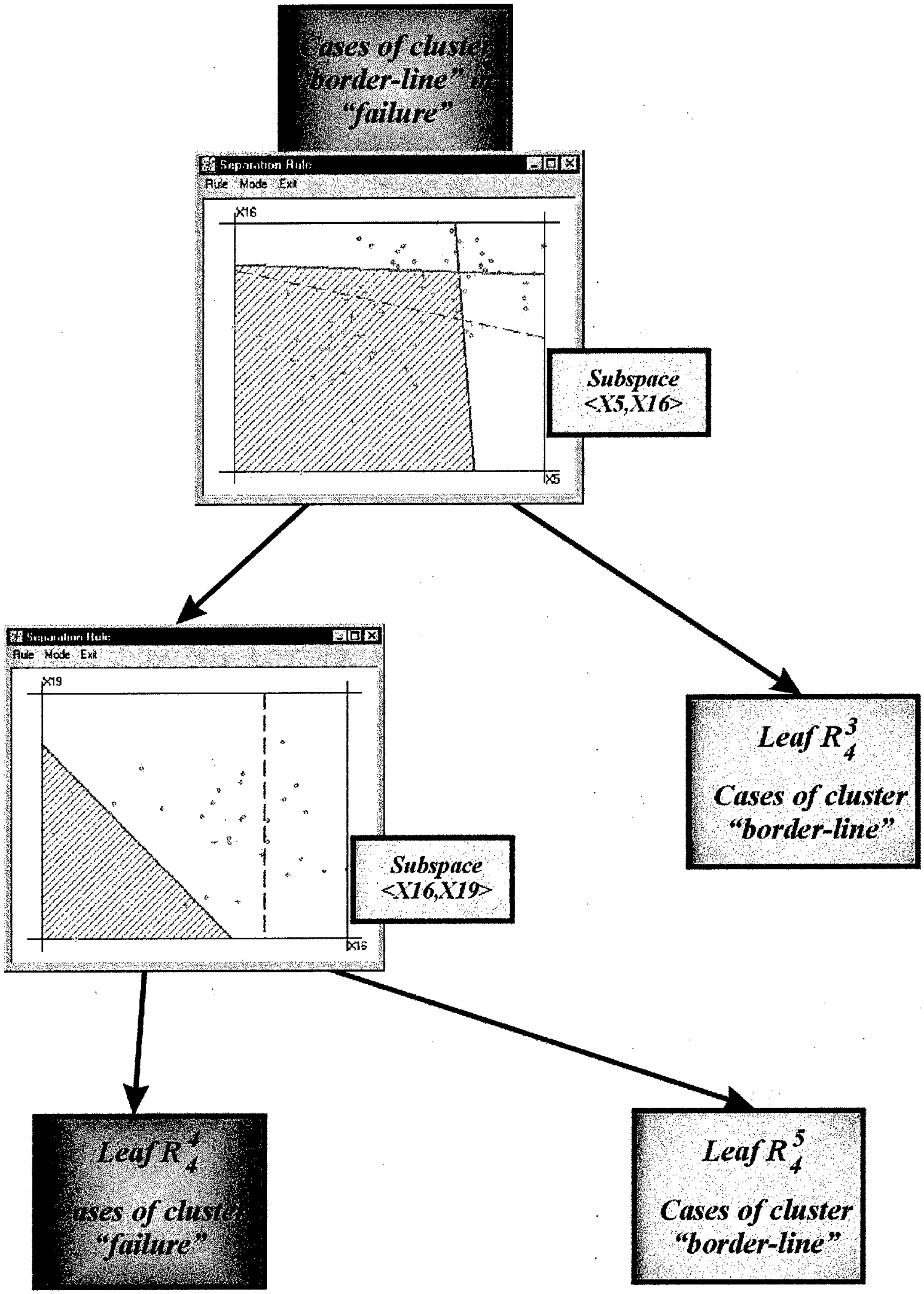


Fig. 4.10. Decision tree #3 for the node of meta-tree <*Status of performance “0” ∨ “1”*>
(see fig. 4.1)

Let us stress again that numerical results given in tab.4.1 were obtained over testing data that were not involved in the learning procedure to design decision making rules.

Table 4.1. Numerical results regarding to the statistical assessment of the quality of decision making procedure over the testing database.

Node of meta-tree	*Probabilit -ies*	*Decision tree fig.4.6.*	*Decision tree fig. 4.8.*	*Decision tree fig. 4.9*	*Decision tree fig.4.10*	*Voting "two of three"*	*Voting "al least one for failure"*
Status of performance "-1"∨"0"∨ "1"	*p(-1/-1)*	0.96					
	p(-1/0∨1)	0.04					
	p(0∨1/-1)	0.05					
	p(0∨1/0∨1)	0.95					
Status of performance "0"∨"1"	*p(0/0)*		0.935	0.91	0.918	0.936	0.952
	p(0/1)		0.065	0.09	0.082	0.064	0.048
	p(1/0)		0.25	0.3	0.19	0.214	0.303
	P(1/1)		0.75	0.7	0.81	0.786	0.697

While designing decision trees we restricted ourselves only by two-level trees. It is obvious that adding one more level could lead to a higher quality of classification up to totally perfect classification. But we aimed at demonstration of benefit of use of collective decision making.

Nevertheless, one can see that resulting classification procedure possess high probabilities of perfect classification. Note that there is no cases of the cluster *"failure"* that was classified as belonging to the cluster *"no failure"*. This result may by considered as an advantage of three–cluster interpretation of the cases of database proposed in this research. Note as well that utilization of collective of decision trees makes it possible to reach more high level of perfect classification. In *Appendix A2* results of testing of each decision tree regarding to each case of testing data and results of use voting procedures according to two schemes are given.

5.1. Algebraic Bayes' Networks for Knowledge Engineering

5.1. Introduction

The theoretical foundation of Algebraic Bayes' Networks (ABN) and its utilization in Knowledge Engineering as a formal model for experts' knowledge formal specification were considered in details in the Interim Report [IR-98]. It was justified that ABN model makes possible to solve a task of formal specification of knowledge under uncertainty and its consistency maintenance. We come back to this problem in this Report. In addition to the material given in the [IR-98] in this Report we present numerical demonstration of the approach and on the basis of the developed case study.

This model is based on probabilistic approach. In traditional probabilistic models, an uncertainty of expert's statement is described by a real number, i.e. by the probabilities of truth of expert's statements. It is well known and it was shown by examples in [IR-98] that in many practical cases expert is not able at all to estimate precisely probabilities associated with statements about dependencies in a domain. At the best case expert is able to determine the lower and the upper bounds of the above probabilities, i.e. to determine so-called interval probability. This case doesn't match "classical" view on the probability theory.

In its nature, interval probabilities don't fix any probabilistic distribution. If multitude of events with assigned interval probabilities is given then even for the case when these events are probabilistically independent, interval probabilities of events don't determine a probability distribution but determine an indefinite class of distributions. This peculiarity reflects the uncertain nature of expert's knowledge about a domain. To cope with interval-valued probabilities, it was proposed a number of approaches. One of them is so-called Algebraic Bayes' Network formal model proposed by author of this Report and considered in detail in [IR-98]. One more source of problems associated with uncertain expert knowledge processing and representation is its inconsistency that in many cases takes place.

In this Report we omitted the main part of conceptual explanation of ABN idea and don't concern to interrelations between ABN model and other formalisms. These aspects were presented in [IR-98] in details. Instead, below we describe the ABN formal model, its components and related algorithms and focused on its application-oriented details. Of course, a part of this section coincides with material given in [IR-98]. The main part of material is repeated to make this section in some sense self-contained and comprehensible in respect to the application-oriented material and numerical demonstrations.

5.2. Properties of expert information

At first phase of life circle of any technical device, as a rule there is no statistical data to develop its diagnostic model. The most frequent case is that only experts' information may be available if there are experts experienced with prototypes or analogues.

Unfortunately human knowledge is often imprecise and inconsistent to the extent to be possible to specify this knowledge in precise formal terms. The latter leads to a number of difficult problems in its practical use within knowledge-based decision support systems, in particular, in diagnostic systems. Human reasoning is very difficult for formal modeling as well. Although experts' knowledge and human reasoning formal specification is a subjects of intensive and deep research during at least last two decades it remains to be a hot problem in Knowledge Engineering up to now.

It was noticed in [IR-98] that the causes of uncertainty and inconsistency of expert information may be as follows:

- Qualitative form of expert statements about dependencies over a set of entities of a domain that makes its unique formalization impossible;
- Very close semantics (sense) of the most part of expert's statements that may be expressed in diverse terms. As a rule, different experts express the same dependency in different words;
- Diversity of experts' experience what can be a cause of hard contradictions.

- Incompleteness of experts' knowledge; in addition, experts are often unable to verbalize their knowledge.

It is generally adopted that to formalize expert's knowledge, an approximate model is to be used. There is no sense to use any precise formal model to formalize imprecise information. It was shown in [IR-98] that expert is not able to estimate probabilities of any statements or combination of statements precisely, i.e. in terms of point-wise probability estimation. In addition, it can be demonstrated by numerical examples that probabilistic models designed on the basis of an expert information specified in terms of point-wise probabilities should be inconsistent inevitably. Therefore, such models cannot be used for correct decision making.

5.3. Advantages of probabilistic model of expert information vs. fuzzy one

There exist a number of approaches for dealing with uncertainty of knowledge. Among them several pseudo-physical logic, many-valued logic, fuzzy logic and theory of possibilities ([Zadeh-78], [Dubois et al -88]) are the most popular ones. But probabilistic model takes a noticeable role because it has a number of important advantages comparing to other mentioned ones. They are as follows:

- probabilistic model is based on a well developed mathematical theory, i.e. probabilistic theory;
- probabilistic models are computationally feasible: due to Large Number theorems, we can deduce and check probabilistic formulae by using repeated statistical simulation of representative size; model output obtained theoretically within probabilistic approach may be validated by simulation;
- between all known formalisms, probabilistic methods provide the most well-developed description of dependency which in probabilistic approach is described by the notion of conditional probability.

Unfortunately, classical probabilistic models based on traditional axiomatic approach are not well fitted for modeling of uncertainty of expert information and human reasoning. The main reason is that it is not clear how probabilities could be obtained. As a rule, empirically assessed probabilities are imprecise and such probabilities can be assigned by confidence interval. A large amount of expert information cannot be expressed in any verbal form. The last deficiency is common for all formal models of uncertainty.

In [IR-98] we have analyzed the existing probability-based approaches developed to deal with interval probabilities ([Dempster-66], [Shafer-76], [Fagin et al-88], [Fagin et al-89]). It was shown that ABN formal model may be considered as a special case of the approach developed in [Fagin et al-88] and [Fagin et al-89]. In this approach a probabilistic space is introduced in axiomatic way via multitude of so-called basic random events that may be dependent in probabilistic sense and may be incomplete. The latter means that in such probabilistic space there exist random events that do not belong to the algebra of event that is generated by the multitude of basic events and operations of union ($\cup$), intersection ($\cap$) and complement (/).

However, Fagin's at al model which uses interval probabilistic measure of knowledge uncertainty is very difficult for implementation in practice, requires a technique to deal with expert's information as it is and doesn't propose a standard way of expert's knowledge representation. ABN model is an attempt to overcome these problems and oriented on practical cases of information extracted from experts.

5.4. Concept of knowledge piece and background probabilistic knowledge

In this subsection the basic concepts and notions to build a model of so-called Algebraic Bayes' Network (ABN) are introduced. This model, in turn, is utilized for expert information formal specification and consistent processing.

5.4.1. Denotations

Let $\Phi_0 = \{x_1, x_2, \ldots, x_n\}$, $x_i \in \{false, true\}$ be a multitude of propositions and $F(X)$= $F(x_{i_1}, x_{i_2}, \ldots, x_{i_k})$ be a formula from the set of well formed formulae given over Φ_0 where X= $\{x_{i_1}, x_{i_2}, \ldots, x_{i_k}\}$– is a subset of propositions Φ_0. Sometimes we shall consider a subset

$X=\{x_{i_1},x_{i_2},\dots,x_{i_k}\}$ as a *tuple* (vector) or as s*equence* of symbols without changing a denotation if its sense will be clear from the context.

Propositions $\{x_{i_1},x_{i_2},\dots,x_{i_k}\}$ can be used with negation (for example, $\neg\ x_i)$ or without it (x_i). To avoid confusion for the case when we use symbol x_i in the sense of argument name (it can be substituted by literal with negation or without it) we denote the last case as $\widetilde{x}_i$ putting symbol "~" above corresponded name. This way is used to denote a set of arguments, therefore $\widetilde{X}$ is denotation of the set of all components of the vector X. For disjunction and conjunction we use symbols *"$\vee$" and "$\wedge$"* respectively. In most cases the symbol *"$\wedge$" is* omitted.

Let $\widetilde{X}=\{\widetilde{x}_{i_1},\widetilde{x}_{i_2},\dots,\widetilde{x}_{i_k}\}$ be a set of propositions. Two subsets $\widetilde{X}_1$ and $\widetilde{X}_2$ such that $\widetilde{X}_1\cup\widetilde{X}_2=\widetilde{X}$, and $\widetilde{X}_1\cap\widetilde{X}_2=\varnothing$ are called partitions of a set $\widetilde{X}=\{\widetilde{x}_{i_1},\widetilde{x}_{i_2},\dots,\widetilde{x}_{i_k}\}$. A symbol of proposition *"$\widetilde{x}_i$"* can be interpreted as denotation of a random event as well. Similar, a conjunction of propositions may be considered as complex random event that is realized if all its arguments (propositions with or without negation) are assigned by truth value *"true"*. In such an interpretation of a conjunction of propositions we may call it as random binary sequence and assign its components a value *"1"* or *"0"* depending on whether the corresponded random event is realized or not.

It should be noted that interpretation of a conjunction of propositions either as formulae or as random event is based on the isomorphism of algebra of logic formulae and algebra of random events that was discussed elsewhere in the literature.

Below in the next subsection we consider basic components of Algebraic Bayes' Network, i.e. so-called "knowledge pieces" of different ranks. In the [IR-98] these notions were discussed in details and were explained conceptually. In this Report we describe them and ANB itself in more formal way. Instead, the practical application is discussed in more depth.

5.4.2. Knowledge Piece of rank 2 (two-propositions knowledge piece)

Let us consider a two-element multitude of propositions $X_{(2)}(i,j)=\{x_1,x_2\}$. Proposition without negation we call as positive one. Let us introduce in a standard way an order relation over the family set over the set $X_{(2)}(i,j)=\{x_i,x_j\}$, i.e. $\aleph=\{\{x_i\},\{x_j\},\{x_i,x_j\}\}$. Each subset of this family set can be mapped to the conjunction constituted by positive propositions contained within it. These propositions may be ordered in the same way as corresponded subsets of the family set $\aleph$. Finally, let each conjunction be assigned a probability.[1]

Definition 5.1. Knowledge piece of rank 2 is a partially ordered set of positive conjunctions that corresponds to elements of family set of the set $X_{(2)}(i,j)$ each assigned a truth probability, i.e.

$$K^{(2)}(x_i,x_j)=\{<\{x_i\},p(x_i)>,<\{x_j\},p(x_j)>,<\{x_i,x_j\},p(x_ix_j)>\}. \tag{5.1}$$

Propositions $\{x_i,x_j\}$ constituting knowledge piece $K^{(2)}(x_i,x_j)$ are called its arguments. ■

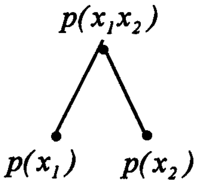


Fig.5.1. Hasse diagram for knowledge piece $X_{(2)}=\{x_1,x_2\}$

Graphical representation of a knowledge piece $K^{(2)}(x_i,x_j)$ of rank 2 in the form of Hasse diagram [Birkhoff-67] is depicted in the fig.5.1.

Definition 5.2. Knowledge piece $K^{(2)}(x_i,x_j)$ is called *consistent* if the following constraints are met:

$$p(x_1)\le 1, \tag{5.2}$$

$$p(x_2)\le 1, \tag{5.3}$$

$$p(x_1)-p(x_1x_2)\ge 0, \tag{5.4}$$

$$p(x_2)-p(x_1x_2)\ge 0, \tag{5.5}$$

[1] Let us emphasise that the probabilities have to meet probabilistic axioms.

$$1-p(x_1)-p(x_2)+p(x_1x_2)\geq 0.\blacksquare \tag{5.6}$$

Inequalities (5.2)-(5.3) correspond to the probabilistic axioms and represent background knowledge of probability theory that has to be met in any case. If these inequalities are held for probabilities $p(x_1)$, $p(x_2)$, $p(x_1x_2)$ then there exist an assignment of probabilities of all other formulae defined over propositions x_i, x_j which probabilities meet probabilistic axioms.

5.4.3. Knowledge piece of rank 3 (three-propositions knowledge piece)

The notion of knowledge piece of the rank 3 is defined like one of the rank 2.

Definition 5.3. Knowledge piece of rank 3 is a partially ordered set of positive conjunctions that corresponds to elements of the family set of the set $X_{(3)}(i,j,k)=\{x_i,x_j,x_k\}$ each assigned a truth probability, i.e.

$$\begin{gathered}K^{(3)}(x_i,x_j,x_k)=\{<\{x_i\},p(x_i)>,<\{x_j\},p(x_j)>,<\{x_k\},p(x_k)>,<\{x_i,x_j\},\\ p(x_ix_j)>,<\{x_i,x_k\},p(x_ix_k)>,<\{x_j,x_k\},p(x_jx_k>,<\{x_i,x_j,x_k\},p(x_ix_jx_k\}>\}.\end{gathered} \tag{5.7}$$

Propositions $\{x_i,x_j,x_k\}$ constituting knowledge piece $K^{(3)}(x_i,x_j,x_k)$ are called its *arguments*. ■

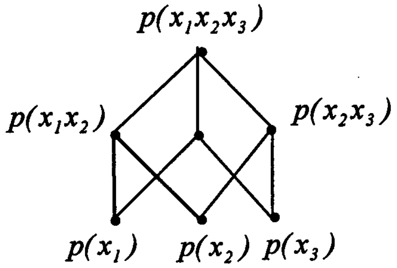


Fig..5.2. Hasse diagram for knowledge piece $X_{(3)}=\{x_1,x_2,x_3\}$

Graphical representation of a knowledge piece $K^{(3)}(x_i,x_j,x_k)$ of rank 3 is depicted in the fig.5.2.

Definition 5.4. Knowledge piece $K^{(3)}(x_i,x_j,x_k)$ is called *consistent* if all knowledge pieces of rank 2 contained in it are consistent according to the *Definition 5.2* (i.e. knowledge pieces $K^{(2)}(x_i,x_j)$, $K^{(2)}(x_i,x_k)$, $K^{(2)}(x_j,x_k)$ are consistent) and in addition the following constraints are met:

$$p(x_jx_k)-p(x_ix_jx_k)\geq 0, \tag{5.8}$$

$$p(x_ix_k)-p(x_ix_jx_k)\geq 0, \tag{5.9}$$

$$p(x_ix_j)-p(x_1x_2x_3)\geq 0, \tag{5.10}$$

$$p(x_k)-p(x_ix_k)-p(x_jx_k)+p(x_ix_jx_k)\geq 0, \tag{5.11}$$

$$p(x_j)-p(x_ix_j)-p(x_jx_k)+p(x_ix_jx_k)\geq 0, \tag{5.12}$$

$$p(x_i)-p(x_ix_j)-p(x_ix_k)+p(x_ix_jx_k)\geq 0, \tag{5.13}$$

$$1-p(x_i)-p(x_j)-p(x_k)+p(x_ix_j)+p(x_ix_k)+p(x_jx_k)-p(x_ix_jx_k)\geq 0.\blacksquare \tag{5.14}$$

Consistency conditions formulated in the *Definition 5.4* for knowledge piece of rank 3 have the same sense how it was commented for consistency conditions of knowledge piece of rank 2. Let us repeat them in brief. If knowledge piece of rank 3 is consistent then there exists a consistent assignment of truth probabilities of all other formulae that may be defined over the same propositions[2]. Inequalities (5.9)-(5.14) represent background knowledge of probability theory regarding to the probabilities associated with knowledge piece of rank 3.

We can define the notion of knowledge piece of any arbitrary rank but no more than the number of atomic propositions in the set Φ_0 (see [IR-98]).

In *Appendix* A3 the full form of consistency conditions are given for the knowledge pieces up to the rank 4.

[2] It is shown below that they can be calculated uniquely.

Fortunately, in practice of knowledge engineering we may restrict ourselves by considering knowledge pieces of rank of not more than 4 (see [IR-98]). On the one hand, it was discovered experimentally that expert is not able to assess dependency over more than three atomic statements. On the other hand, if a rank of knowledge piece is increased then reliability of the involved probability assessment decreases significantly. Therefore we may restrict ourselves by considering of knowledge pieces of the rank no more than 3. Below it is shown that for consistency maintenance we need to deal with knowledge pieces of rank 4 ([IR-98)].

5.5. Algebraic Bayes' Network: Formal Definition

The above materials form a foundation for introduction a notion of the Algebraic Bayes' Network that is basic one for design of algorithms of expert information processing. In this subsection the corresponding formal framework is described.

It was mentioned above that experts are able to talk reliably about dependencies over no more then three atomic statements, i.e. about truth probabilities of formulae determined over sets of propositions $X_{(1)}(i)=\{x_i\}$, $X_{(2)}(i,j)=\{x_i,x_j\}$ and $X_{(3)}(i,j,k)=\{x_i,x_j,x_k\}$. This experimentally inferred conclusion makes possible to restrict ourselves by knowledge pieces of rank 2 and 3 as standard patterns of knowledge. The latter can serve as justification for assumption that after processing of fragments of expert information aimed at consistency maintenance the result may be represented as a set of consistent knowledge pieces as follows:

$$KB=\{\{K^{(1)}(x_i)\}_{i\in I_n},\ \{K^{(2)}(x_i,x_j)\}\ddot{I}_{i,j\in I_n}\},\ \{K^{(3)}(x_i,x_j,x_k)_{i,j,k\in I_n}\}\}. \tag{5.15}$$

However, different instances of knowledge pieces in (5.15) may contain the same propositions and, hence, may be dependent. It means that they have to be structured in a formal framework in such a way that it should be possible to check and maintain consistency of the entire multitude of knowledge pieces. For example, the same conjunctions can be contained in a number of knowledge pieces. Hence, they have to be assigned the same intervals of truth probability but the latter may be not held in for knowledge pieces contained in (5.15) because they might be extracted from different experts and/or independently on each other. This means that a multitude of above knowledge pieces may be inconsistent and, hence, doesn't represent a knowledge base. This multitude is a "half-finished product" only and further processing is needed to constitute knowledge base.

As a structure for joint representation of a multitude of knowledge pieces that makes it possible to detect and eliminate contradictions of knowledge pieces we propose to use a so-called "Algebraic Bayes' Network" (ABN) structure ([Gorodetski-92], [Gorodetski et al-97]).

Definition 5.5. Let us call an Algebraic Bayes' Network a set of consistent knowledge pieces

$$KB^{3}=\{\{K^{(1)}(x_i)\}_{i\in I_n},\ \{K^{(2)}(x_i,x_j)\}\ddot{I}_{i,j\in I_n}\},\ \{K^{(3)}(x_i,x_j,x_k)_{i,j,k\in I_n}\}\}$$

structured as Hasse diagram (semi-lattice) ([Birkhof-85]) added with

- a multitude of constraints conditioned by probabilistic axioms over truth probabilities of all conjunctions that are contained in the above multitude *KB*; (they form so-called background knowledge);
- algebra of formulae; and
- inclusion–exclusion formulae for truth probabilities of formulae of the multitude *KB*. ■

Let us explain the introduced notion by example.

Example 5.5.1.

Let ABN consist of four knowledge pieces formed due to expert information about possible dependencies over six propositions $\{x_1,x_2,\dots,x_6\}$ ([Gorodetski et al-97]):

$$KB=\{K^{(3)}(x_1,x_2,x_3),\ K^{(3)}(x_2,x_3,x_4),\ K^{(2)}(x_4,x_5),\ K^{(2)}(x_5,x_6)\}.$$

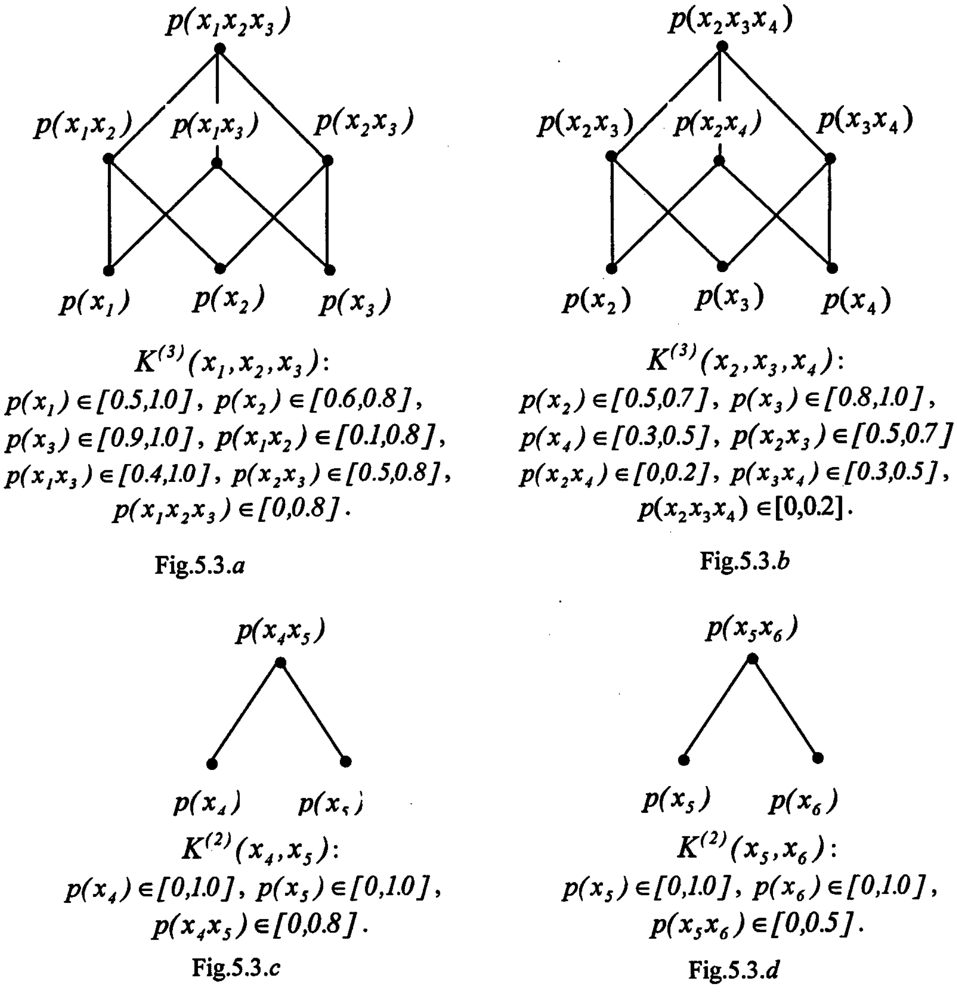
[3] Denotation *"KB"* for this set emphasizes the fact that it can be considered as knowledge base.

Graphical presentation of these knowledge pieces is given in the fig. 5.3. In the fig.5.4 graphical representation of the corresponding (structured) ABN is given.

It can be shown that ABN model is a special case of model known as extended probabilistic space ([Fagin et al-88]). Indeed, ABN is an extended probabilistic space under the following assumptions:

1. A multitude of basic events (algebra of family set) isomorphous to the algebra of logic formulae is not known. Interval assessments of truth probabilities are known for a subset of logic formulae over set of atomic propositions $\Phi_0 = \{x_1, x_2, \dots, x_n\}$;
2. ABN contains only positive conjunctions of atomic propositions of a length no more than 3.

The necessity of both above assumptions is conditioned by a specific of application that has to deal with expert information processing. Indeed, the first assumption is conditioned by a limit of information that can be extracted from experts, and the second one reflects the restricted possibility of an expert to discover dependencies. It should be noted that a model of extended probabilistic space ([Fagin et al-88]) requires too much information to be practically helpful within a knowledge engineering tasks.



$K^{(3)}(x_1,x_2,x_3)$:
$p(x_1)\in[0.5,1.0]$, $p(x_2)\in[0.6,0.8]$,
$p(x_3)\in[0.9,1.0]$, $p(x_1x_2)\in[0.1,0.8]$,
$p(x_1x_3)\in[0.4,1.0]$, $p(x_2x_3)\in[0.5,0.8]$,
$p(x_1x_2x_3)\in[0,0.8]$.

Fig.5.3.*a*

$K^{(3)}(x_2,x_3,x_4)$:
$p(x_2)\in[0.5,0.7]$, $p(x_3)\in[0.8,1.0]$,
$p(x_4)\in[0.3,0.5]$, $p(x_2x_3)\in[0.5,0.7]$,
$p(x_2x_4)\in[0,0.2]$, $p(x_3x_4)\in[0.3,0.5]$,
$p(x_2x_3x_4)\in[0,0.2]$.

Fig.5.3.*b*

$K^{(2)}(x_4,x_5)$:
$p(x_4)\in[0,1.0]$, $p(x_5)\in[0,1.0]$,
$p(x_4x_5)\in[0,0.8]$.

Fig.5.3.*c*

$K^{(2)}(x_5,x_6)$:
$p(x_5)\in[0,1.0]$, $p(x_6)\in[0,1.0]$,
$p(x_5x_6)\in[0,0.5]$.

Fig.5.3.*d*

Fig.5.3. A multitude of knowledge pieces with assigned truth probabilities of the formulae corresponded to its nodes (components of ABN) given below in fig.5.4.

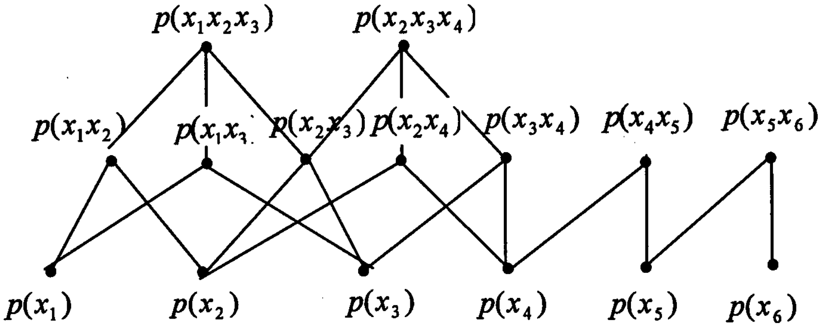


Fig.5.4. An example of graphical representation of ABN contained four knowledge pieces.

5.6. Consistency of Algebraic Bayes' Networks: Background Knowledge

5.6.1. Internal Consistency of Algebraic Bayes' Network

In *Definition 5.5* ABN was determined as structured multitude of consistent knowledge pieces. It can be shown that consistency of each knowledge piece does not guarantee consistency of ABN in strict sense. Let n be a cardinality of the multitude of propositions $\Phi_0 = \{x_1, x_2, \dots, x_n\}$ that are arguments of ABN. To check consistency of such ABN in the strict probabilistic sense, it should be necessary to form a knowledge piece of rank n and to check its consistency in the same way which was specified above for knowledge pieces of rank *2* and *3*. But ABN of such rank consists of $(2^n - 1)$[4] elements and probabilistic background knowledge imposes $3^n - 1$ equations or inequalities. Even for relatively low value of n the tasks of checking consistency (checking the existence of a solution of the corresponding inequalities) is too complex.. It is clear that such size of constraint satisfaction task is too high to be used in practice. However, dimensions of real life applications may be much more. This means that "direct way" of consistency checking and maintaining is infeasible.

On the other hand, as a rule, expert's probabilities are very approximate and incomplete and this is a justification for use of more weak condition for checking and maintaining consistency. To assess consistency, we propose to utilize a notion of linearly ordered set of ABN consistency degrees.

Definition 5.6. Algebraic Bayes' Network is *locally consistent* if and only if all its knowledge pieces are consistent. ■

Local consistency of ABN is the minimal (the weakest) degree of its consistency. I should be noted that local consistency is provided by *Definition 5.5* of ABN. It is obvious that local consistency is a necessary consistency condition of ABN.

It was mentioned above that different knowledge pieces may contain common conjunctions. For example, ABN depicted in the fig.5.4 contain conjunctions x_2, x_3, x_2x_3, x_4, x_5, x_6 that are included in several knowledge pieces. To be locally consistent, such conjunctions can be assigned by distinct values of truth probabilities in different knowledge pieces because experts can provide a contradictory information and different knowledge pieces can be formed by experts of different experience. Within ABN common conjunctions have to be equal.

Definition 5.7. Two knowledge pieces of an ABN are called *coordinated* if truth probabilities of all common conjunctions within these knowledge pieces are assigned equal values. ■.

Definition 5.8. Locally consistent ABN is called *internally consistent* if and only if all its knowledge pieces are coordinated. ■

[4] This number is equal to the total number of conjunctions of length no more then n without empty conjunction.

Algorithm of checking and maintenance of ABN internal consistency is relatively simple. Let us describe it idea.

Let all knowledge pieces be enumerated by indexes from 1 to M, where M is a total number of knowledge pieces in ABN. Let Ω be a multitude of all knowledge pieces. Let us denote each knowledge piece number i by KP_i, $i \in 1,2,...,M$.

Algorithm of maintenance of internal consistency of ABN:

1. In the multitude Ω select a knowledge piece KP_i indexed by minimal number i and then select all knowledge pieces having indexes $k>i$, $k \leq M$ and constituting nonempty multitude of common conjunctions together with KP_i. Let us denote the multitude of such knowledge pieces including KP_i as $\boldsymbol{KP}_{\geq i}$;
2. Form a list of common conjunctions $\boldsymbol{Con}_{\geq i}$ of the multitude $\boldsymbol{KP}_{\geq i}$ and for each conjunction $C_j \in \boldsymbol{Con}$ compute intersections of intervals of truth probability assigned to each conjunction within all knowledge pieces of the multitude $\boldsymbol{KP}_{\geq i}$ (This operation intend to make all knowledge pieces coordinated). If all above intersections are nonempty then assign the resulted (coordinated) intervals to the corresponding conjunctions. Otherwise, ABN is internally inconsistent;
3. Delete knowledge piece KP_i from the multitude Ω. If Ω is nonempty then go to the step 2;.
4. Solve tasks of local consistency maintenance for each KP_i that contains at least one interval of truth probability assigned a new value in the step 2. ■

After running steps 1 - 4 the possible outputs can be as follows:

- either *inconsistency* of ABN is ascertained;
- either coordination of all knowledge pieces is ascertained and therefore ABN *internal consistency* is reached;
- or not all knowledge pieces turns out coordinated. In the last case all steps 1 - 4 have to be repeated once more.

The above algorithm converge to the decision under search because all intervals are changed monotonically (they may become only more narrow) and value of intervals are restricted from below by empty set.

Example 5.6.1.

Let us demonstrate the algorithm of internal consistency maintenance for ABN depicted in the fig.5.4 assigned by truth values given in the fig.5.3 ([Gorodetski et al- 97]).

The first run through the steps 1-3.

Consider the first pair of knowledge pieces, i.e. $K^{(3)}(x_1,x_2x_3)$ and $K^{(3)}(x_2,x_3x_4)$. They have three common conjunctions but only one is needed to be coordinated. The result is as follows:

$$p^*(x_2) \in [0.6,0.7].$$

Two other common conjunctions preserve their truth probability intervals, i.e.

$$p^*(x_3) \in [0.9,1.0],\ p^*(x_2x_3) \in [0.5,0.7].$$

The second run through the steps 1-3.

Consider the second pair of knowledge pieces, i.e. $K^{(3)}(x_2,x_3x_4)$ и $K^{(2)}(x_4,x_5)$. For them one interval of truth probability is modified and takes value

$$p^*(x_4) \in [0.3,\ 0.5].$$

The third run through the steps 1-3.

Consider the pair of knowledge pieces $K^{(2)}(x_4,x_5)$ и $K^{(2)}(x_5,x_6)$. Their common conjunction x_5 is coordinated. Resulting value of its interval of truth probability is preserved:

$$p(x_5)\in[0,1.0].$$

Run of the step 4.

Run procedures of local consistency maintenance for all knowledge pieces of ABN. The final result is as follows ([Gorodetski et al-97]):

$K^{(3)}(x_1,x_2,x_3)$:
$p(x_1)\in[0.5,1.0]$, $p(x_2)\in[0.6,0.7]$,
$p(x_3)\in[0.9,1.0]$, $p(x_1x_2)\in[0.1,0.7]$,
$p(x_1x_3)\in[0.4,1.0]$, $p(x_2x_3)\in[0.5,0.7]$,
$p(x_1x_2x_3)\in[0,0.7]$.

$K^{(3)}(x_2,x_3,x_4)$
$p(x_2)\in[0.6,0.7]$, $p(x_3)\in[0.9,1.0]$,
$p(x_4)\in[0.3,0.5]$, $p(x_2x_3)\in[0.5,0.7]$,
$p(x_2x_4)\in[0,0.2]$, $p(x_3x_4)\in[0.3,0.5]$,
$p(x_2x_3x_4)\in[0,0.2]$.

$K^{(2)}(x_4,x_5)$:
$p(x_4)\in[0.3,0.5]$, $p(x_5)\in[0.5,0.7]$,
$p(x_4x_5)\in[0,0.5]$.

$K^{(2)}(x_5,x_6)$
$p(x_5)\in[0.5,0.7]$, $p(x_6)\in[0.3,0.5]$,
$p(x_5x_6)\in[0,0.5]$.

After running the step 4 of algorithm all conjunctions are assigned by coordinated values of truth probabilities. Therefore, the resulting ABN is internally consistent. ■

The above algorithm does not cover the case when there exist a number of order relations over truth probabilities of conjunctions. This case was considered in [IR-98] and we demonstrate it numerically below within the case study related to application of ABN to the carburetor of a car engine diagnosis in S*ubsection 5.7*.

Unfortunately, internal consistency of ABN doesn't guarantee the strict consistency of the latter. Internal consistency condition is only necessary but not sufficient. Let us consider example demonstrating the case when ABN is internally consistent but not consistent in the strict sense.

Example 5.6.2.

Let us consider ABN depicted in the fig.5.5 that consists of three knowledge pieces of rank 2 ([Gorodetski et al-97]) and which conjunctions are assigned truth value intervals as follows:

$p(x_1)\in[0.5,1.0]$, $p(x_2)=0.5$, $p(x_3)=0.5$, $p(x_1x_2)=0.5$, $p(x_1x_3)=0.5$, $p(x_2x_3)=0$.

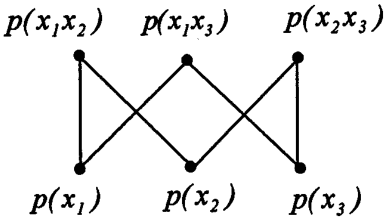


Fig.5.5. An example of internally consistent ABN that is inconsistent in strict sense

It can be shown that this ABN is inconsistent. In accordance with *Definition 5.10* this ABN is internally consistent. However, in this ABN there does not exists any consistent assignment for truth probability of conjunction $x_1x_2x_3$ that meets the probabilistic axioms imposed on this conjunction, i.e. interval of truth probability $p(x_1x_2x_3)$ of this ABN is empty. ■

5.6.2. External consistency

Let us introduce one more degree of ABN consistency that corresponds to the next degree of consistency compared to the both local and internal degrees of ABN consistency.

Definition 5.10. Internally consistent ABN is called *externally consistent* if assignment of truth probabilities of all its maximal elements is consistent. ■

Let us first explain new terms used in the *Definition 5.10.* In *Definition 5.10* a notion of *maximal element* of ABN is introduced. This notion is used in the lattice theory sense. A Hasse diagram ([Birkhof-67]) of ABN that is used for graphical representation of the latter is a semi-lattice and, saying informally, an element of ABN is the maximal one if it corresponds to a conjunction of maximal length within a knowledge piece to which it belongs. It should be noted that a maximal element can belong to an only knowledge piece and each knowledge piece may contain only maximal element ([Birkhof-67]). For example, in ABN depicted in 5.4 maximal elements are $x_1x_2x_3$, $x_2x_3x_4$, x_4x_5 and x_5x_6. It is obvious that the maximal element of a knowledge piece determine uniquely the *structure* of the latter.

Let us now explain the term "*externally consistent* assignments of truth probabilities of all maximal elements". Each maximal element and corresponding knowledge piece may be considered as contained in a knowledge piece of a higher rank. For example, knowledge pieces $K^{(2)}(x_1,x_2)$, $K^{(2)}(x_1,x_3)$ and $K^{(2)}(x_2,x_3)$ that form ABN depicted on the fig.5.4 may be considered as contained within knowledge piece $K^{(3)}(x_1,x_2,x_3)$ of rank 3 depicted in the fig.5.5. Knowledge pieces depicted in the fig.5.3*a*, 5.3.*b* may be considered as contained within the knowledge piece of rank 4 depicted in the fig.5.7. In the both above mentioned figures dotted lines correspond to the edges of Hasse diagram that are not contained in the original knowledge pieces.

Let us consider an example to demonstrate the basic idea of notion of external consistency.

Example 5.6.3.

Let us consider ABN containing a multitude of knowledge pieces depicted in the fig. 5.5 and marked by unbroken lines. This ABN is internally consistent and truth probabilities of its conjunctions are as follows:

$$p(x_1)\in[0.5,0.6],\ p(x_2)\in[0.5,0.6],\ p(x_3)\in[0.5,0.6],$$
$$p(x_1x_2)\in[0.5,0.6],\ p(x_1x_3)\in[0,0.5],\ p(x_2x_3)\in[0.4,0.6]. \qquad (5.16)$$

In the original ABN a truth probability of the conjunction $x_1x_2x_3$ is absent but it has to meet constraints imposed by background probabilistic knowledge (5.2) - (5.6) and (5.8)-(5.14). It should be noted that a subset of these constraints has been already met in the original ABN because the latter is internally consistent. If intervals of some truth probabilities were narrowed, then, nevertheless, already met constraints would not be broken. For example (see fig. 5.6) the probabilities $p(x_1x_2)$, $p(x_1x_3)$, $p(x_2x_3)$ are consistent internally within corresponding knowledge pieces of rank 2. To meet the external consistency conditions, the following constraints have been held:

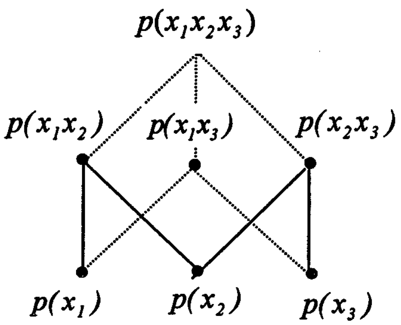


Fig.5.6.Knowledge piece of rank 3 that contains maximal elements of knowledge pieces of rank 2

$$p(x_2x_3)-p(x_1x_2x_3)\geq 0, \qquad (5.17)$$
$$p(x_1x_3)-p(x_1x_2x_3)\geq 0, \qquad (5.18)$$
$$p(x_1x_2)-p(x_1x_2x_3)\geq 0, \qquad (5.19)$$
$$p(x_3)-p(x_1x_3)-p(x_2x_3)+p(x_1x_2x_3)\geq 0, \qquad (5.20)$$
$$p(x_2)-p(x_1x_2)-p(x_2x_3)+p(x_1x_2x_3)\geq 0, \qquad (5.21)$$
$$p(x_1)-p(x_1x_2)-p(x_1x_3)+p(x_1x_2x_3)\geq 0, \qquad (5.22)$$
$$1-p(x_1)-p(x_2)-p(x_3)+p(x_1x_2)+p(x_1x_3)+$$
$$+p(x_2x_3)-p(x_1x_2x_3)\geq 0. \qquad (5.23)$$

This task is simple enough. It results in externally consistent solution that differs from the original internally consistent ABN only in the estimation of truth probability of the conjunction x_1x_3 that is $p(x_1x_3) \in [0.3,0.5]$.■

Thus, a task of checking and maintenance of external consistency of an ABN is reduced to the checking and maintenance consistency of a knowledge piece of minimal rank that contains corresponding maximal elements of ABN.

Let us consider the task in a more formal way.

Definition 5.12. Knowledge piece of minimal rank that contains maximal elements of a given multitude of knowledge pieces $\boldsymbol{K}=\{K^{(i)}\}_{i=1}^{q}$ is called minimal external knowledge piece for the multitude $\boldsymbol{K}$.■

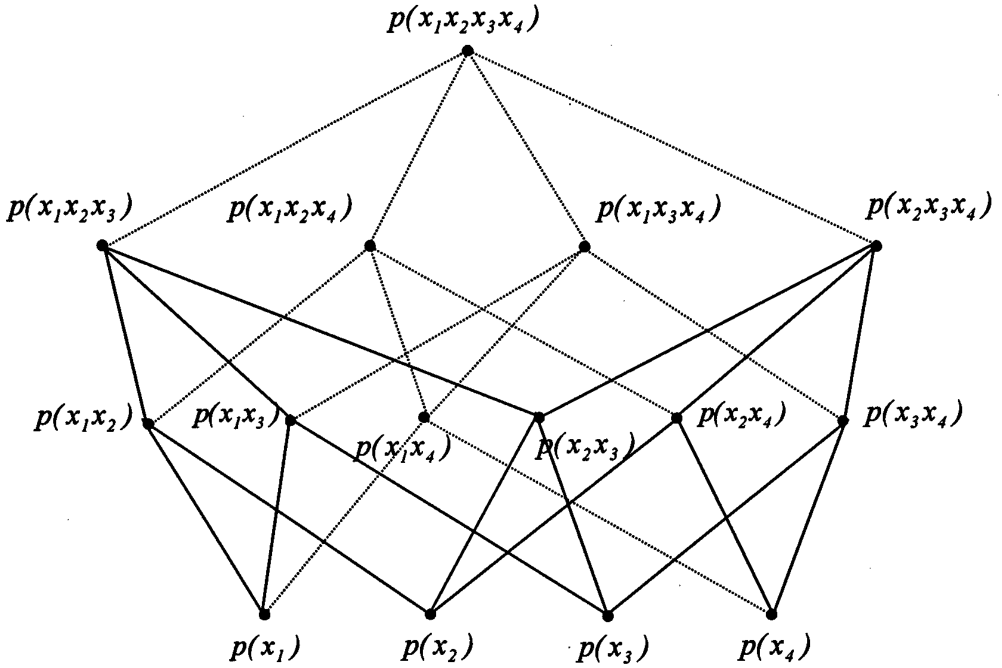


Fig.5.7. Knowledge piece of rank 4 that contains maximal elements of knowledge pieces of rank 3

In fig.5.6 and fig5.7 the minimal external knowledge pieces of rank 3 and 4 respectively are depicted.

Let us remind that we consider ABN that contains knowledge pieces of rank of no more then 3. Therefore, we can restrict ourselves by the following particular cases:

1. External consistency of a subset of knowledge pieces of rank 2 contained in a minimal external knowledge piece of rank 3. It is clear that every minimal external knowledge piece may contain no more than 3 knowledge pieces of rank 2;
2. External consistency of a subset of knowledge pieces of rank 3 contained in a minimal external knowledge piece of rank 4. In this case a minimal external knowledge piece may contain two, three or four knowledge pieces of rank 3;
3. External consistency of a subset of knowledge pieces that contains knowledge pieces either rank 2 or rank 3. The number of variants of knowledge pieces in this case is more than in the previous two ones.

Let us consider above listed special cases.

1.External consistency of ABN that contains knowledge pieces of ranks no more than 2

In this case the task is reduced to one that deals only with internal consistency of knowledge pieces of rank 3. Algorithm of the task solving was demonstrated in the example 5.6.3 and consists in meeting constraints (5.17) - (5.23). In principal, the set of constraints (5.17)– (5.23) can be simplified if there is no necessity to estimate intervals of truth probabilities of conjunctions that are not contained in the original ABN. For example, if ABN consists of two knowledge pieces, say, $K^{(2)}(x_1,x_2)$, $K^{(2)}(x_1,x_3)$ (see fig.5.6), then there is no need to estimate probabilities $p(x_2x_3)$ и $p(x_1x_2x_3)$. This task is slightly simpler, however, in such approach we do not preserve a standard form of task formulation. To preserve it, one has to take into account the full multitude of constraints.

Figures 5.6 and 5.7 cover all special cases of ABN that contains knowledge pieces of rank of no more than 2.

2.External consistency of ABN that contains knowledge pieces of rank 3

Fig.5.7 makes possible to enumerate all cases that may be met while dealing with external consistency problem solving for ABN that contains knowledge pieces of ranks no more than 3. An original ABN can contain any pairs, triples of knowledge pieces of rank 3. But it is a reason to preserve the standard formulation of a the set of constraints and does not distinct particular cases.

Thus, for general case of the task under consideration it is necessary to use constraints of the multitude $E^{(4)}$ which are presented in *Appendix A3* of the Report.

3. External consistency of a subset of knowledge pieces that contains a mixture of knowledge pieces of rank 2 and 3.

This case is reduced to the two previous ones. It is clear that if ABN contains at least one knowledge piece of rank 3 then a minimal external knowledge piece is of rank 4 what allows us to reduce this case to the task considered above.

5.7. Case study of Experts' Information Processing: Car Engine Diagnostics

In previous subsections of this section we considered theoretical basic of Algebraic Bayes' Networks (ABN). This structure was developed to be used as a formal framework to deal with expert's information to design knowledge and to integrate together experts' knowledge and other one which obtained via Data Mining and Knowledge Discovery from statistical databases. Peculiarity of expert's information and problem of its use for development of integrated knowledge base were considered in detail in Interim Report [IR-98]. In addition the problem was outlined in this Report in *Subsection 5.2.* Below we describe in brief and demonstrate technology of expert's information processing on the basis of the ABN framework.

Let $\Phi_0 = \{x_1,x_2,...,x_n\}$, $x_i \in \{false,true\}$ be a multitude of propositions that formalize domain factors, for example, those that determine status of performance of a device and $F(X)=F(x_{i_1},x_{i_2},...,x_{i_k})$ be a logic formula from the set of well formed formulae given over Φ_0 where $X=\{x_{i_1},x_{i_2},...,x_{i_k}\}$ — is a subset of propositions Φ_0. It was justified in [IR-98] that if we intend to extract knowledge from an expert we have to take into account that expert is able to reply only on relatively simple and clear questions about dependencies existing in a domain. The dependencies he/she is able to talk about are usually relations given over no more than two or at best three variables. If we intend to use ABN as the formal framework for uncertain experts' knowledge processing and representation the above questions are about experts' assessment of probabilities of occurrence of those or others events given status of device performance, say, *"no failure"* and *"failure"*.

ABN is partially ordered set of positive conjunctions of propositions from the set Φ_0 assigned by probabilities of truth. But in practice experts are able to talk not only about probabilities of positive conjunctions of propositions (complex random events). Instead, they may possess knowledge about probabilities of conjunction containing literals with and without negation, about probabilities of disjunctions that, in its turn, may comprise of positive and negative literals. For example, if we ask an expert about eventual probabilistic dependencies existing over two statements (factors, features), we may await replies about the following probabilities [IR-98]:

$$p(f_1)=p(x_1),\ p(f_2)=p(x_2),\ p(f_3)=p(\neg x_1),\ p(f_4)=p(\neg x_2),$$
$$p(f_5)=p(x_1x_2),\ p(f_6)=p(\neg x_1x_2),\ p(f_7)=p(x_1\neg x_2),$$
$$p(f_8)=p(\neg x_1\neg x_2),\ p(f_9)=p(x_1\vee x_2),\ p(f_{10})=p(\neg x_1\vee x_2),$$
$$p(f_{11})=p(x_1\vee\neg x_2),\ p(f_{12})=p(\neg x_1\vee\neg x_2).$$

It is clear what probabilities we may deal with for the case of three–variables dependencies.

Therefore, one of the problem of expert's information processing is how to obtain ABN model if information consists of a number of experts' statements about probabilities of formulae given over arbitrary two–and three– proposition logic formulae. The corresponding equations and algorithms were considered in [IR-98]. In *Appendix 3* of this Report they are given again. Below we outline corresponding algorithms and demonstrate technology of experts' information processing aimed at ANB consistent formal model design.

5.7.1. Experts' Information: Car engine carburetor diagnostics.

Let we intend to design ABN aims at solving task of carburetor of a car engine diagnostics. We suppose that carburetor status can take two values: *"no failure"* and *"failure"*. Below in the table 5.1 in the column #2 the denotation and physical sense of factors observed by driver are given. They are given as the statements of natural language in quotation marks. Symbol of negation "¬" reflects the fact that corresponding statement in quotation marks is false. Columns #3 and #4 corresponds to the experts' assessment of the respective interval valued probabilities of truth of the statements given in the column #2 for status of carburetor *"failure"* and *"no failure"* respectively.

Table 5.1. Experts' information about subject domain "Car Engine Diagnostics: Carburetor".

No	*Denotation of experts' expressions on natural language*	*p (f /" failure =p(f/1)*	*p(f /"no failurẽ") =p(f/0)*
1	x_1: *"Choking of cold or warmed engine"*	[0.2–0.6]	[0.0–0.3]
2	$\neg x_1$: ¬(*"Choking of cold or warmed engine"*)	[0.4–0.8]	
3	x_2: *"Engine shut down before reaching nominal temperature"*	[0–0.3]	
4	$\neg x_2$: ¬(*"Engine shut down before reaching nominal temperature"*)	[0.6–1.0]	
5	x_3: *"Difficult firing of warmed up engine"*	[0.4–0.8]	[0.2–0.4]
6	$\neg x_3$: ¬(*"Difficult firing of warmed up engine"*)	[0.2–0.6]	
7	x_4: *" Unsteady engine run at no load"*	[0.2–0.6]	[0.0–0.05]
8	$\neg x_4$: ¬(*"Unsteady engine run at no load*")	[0.3–0.7]	
9	x_5: *"Jerking movement at constant speed*'	[0.1–0.4]	[0.0–0.1]
10	$\neg x_5$: ¬(*"Jerking movement at constant speed"*)	[0.6–0.9]	
11	x_6: *"Jerking acceleration"*	[0.2–0.5]	[0.1–0.3]
12	$\neg x_6$: ¬(*"Jerking acceleration"*)	[0.5–0.8]	
13	x_7: *"Low acceleration at movement"*	[0.6–0.9]	[0.1–0.3]
14	$\neg x_7$: ¬(*"Low acceleration at movement"*)	[0.1–0.4]	[0.2–0.6]

No	*Denotation of experts' expressions on natural language*	$p(f /$" *failure* $=p(f/1)$	$p(f /$"*no failure*") $=p(f/0)$
15	x_8 : *"Highest level of engine power is not achievable. Engine troubles at highest load"*	[0.5–0.8]	
16	$\neg x_8$: *¬("Highest level of engine power is not achievable. Engine troubles at highest load").*	[0.2–0.5]	
17	x_9 : *"Running of engine after shut down"*	[0.0–0.3]	
18	$\neg x_9$: *¬('Running of engine after shut down")*	[0.7–1.0]]	
19	x_{10} : *Sputtering engine*	[0.0–0.3]	
20	$\neg x_{10}$:¬"*(Sputtering engine")*	[0.7–1.0]	
21	x_{11} : *Abnormally high level of fuel consumption*	[0.6–0.9]	
22	$\neg x_{11}$: *¬("Abnormally high level of fuel consumption")*	[0.1–0.4]	
23	$x_8\ x_{11}$	[0.4–0.8]	
24	$x_6\ x_8$	[0.1–0.5]	
25	$x_1\ x_3$	[0.1–0.5]	
26	$x_7\ x_8\ x_{11}$	[0.2–0.5]	
27	$x_7\ x_8$	[0.4–0.8]	
28	$x_7\ x_{11}$	[0.4–0.8]	
29	$x_6\ x_7\ x_8$	[0.0–0.4]	
30	$x_6\ x_7$	[0.1–0.6]	
31	$x_6\ x_7\ x_{11}$	[0.1–0.4]	
32	$x_6\ x_{11}$	[0.1–0.5]	
33	$x_1\ x_4$	[0.2–0.6]	
34	$x_3\ x_5$	[0.1–0.6]	
35	$x_3\ x_{11}$	[0.3–0.7]	

Remark: Empty positions of the tab.5.1 correspond to the probability interval [0,1].

In addition to the data given in tab.5.1 the following order relations given over probabilities of truth of the formulae were extracted from experts:

$$\begin{aligned} &p(X_7) \geq 2 \times p(X_9), \\ &5 \times p(X_{10} X_{11}) \geq p(X_4). \end{aligned} \tag{5.24}$$

Data in tab.5.1 and order relation (5.24) forms experts' information that has to be taken into account for knowledge base design.

Let us note that below we demonstrate technology of ABN design and its consistency maintenance only for the experts' information related to the status *"failure"* of device "Carburetor", because this case is much more complex than the case related to the status *"no failure"*.

The demonstrated below technology of experts' information processing to design knowledge base and corresponding numerical results are obtained on the basis of the developed software. The software was developed in the environment Visual C++ and Access 97 Data Base Management System.

5.7.2. Consistency maintenance Procedure: Numerical Results

The first step of processing of the data given in the tab.5.1 is *knowledge pieces extraction.* Printout of this step result is depicted in the fig.5.8. In addition the list of the knowledge pieces of rank 1, 2 and 3 is given in tab.5.2 in the second and third columns. It can be seen that experts' information leads to the Algebraic Bayes' Network comprising 10 knowledge pieces.

Next step according to the developed technology corresponds to the ABN local consistency maintenance that is calculation of the locally consistent probabilities assigned to the nodes of the ABN. To realize this step it is necessary to use equations given in *Appendix A3.* These equations are labeled as $E_1^{(2)} - E_9^{(2)}$ for knowledge pieces of rank 2 and as $E_1^{(3)} - E_{39}^{(3)}$ for knowledge piece of rank 3. Of course, it is necessary to use only those of them that correspond to the experts' information containing in the tab.5.1. One can see that in our case there is no necessity to use equations containing disjunctions of propositions.

Equations $E_1^{(2)} - E_9^{(2)}$ and $E_1^{(3)} - E_{39}^{(3)}$ have to be considered as constraints for pairs of linear programming tasks formulated for each node of the corresponding knowledge piece. This procedure was described in detail in Interim Report [IR-98]. The results of calculation of the truth probabilities of ABN nodes are given in the fifth column of tab.5.2. It should be noted that resulting ABN is locally consistent.

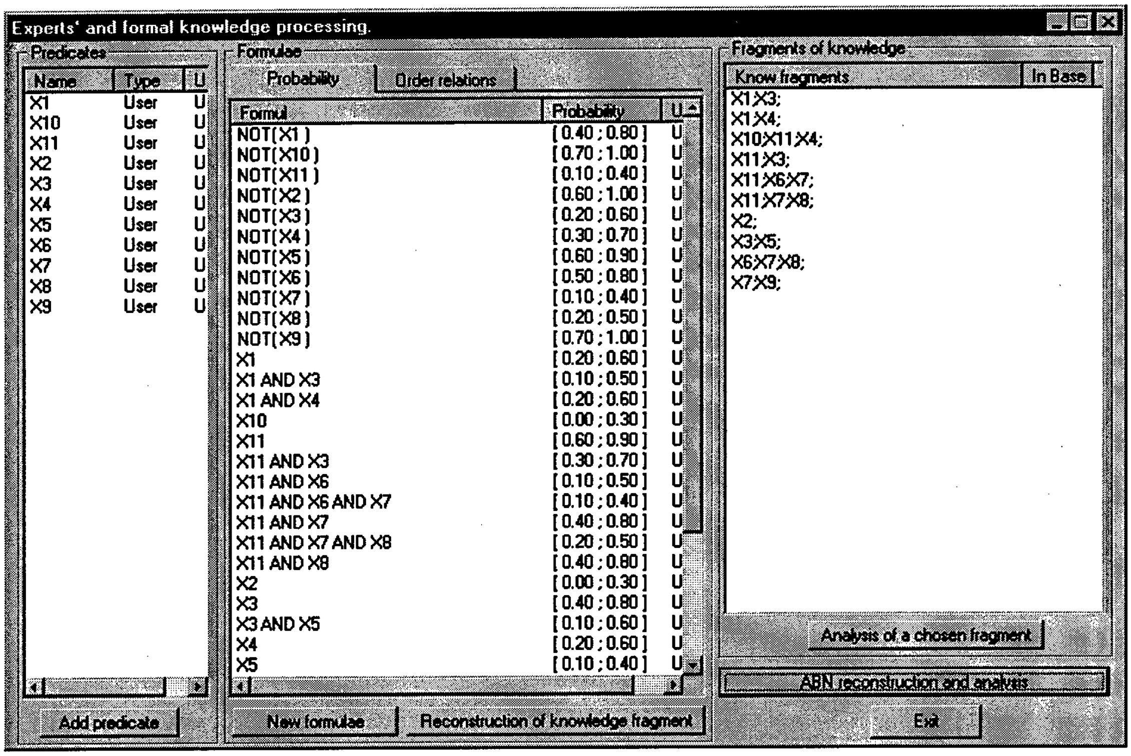
The third step is maintenance of the ABN *internal consistency.* The algorithm of this task solution was described above in *Subsection 5.6.* Let us remind that the algorithm aims at calculating of the coordinated probabilities for the nodes which belong to several (more than to one) knowledge pieces. The results of calculations are presented in the fifth column of tab.5.2.

The last step is checking and if necessary maintenance of the external consistency. This procedure was described in Interim Report [IR-98] and outlined in *Subsection 5.6* above. Results of implementation of the procedure are given in the sixth column of tab.5.2.

Table 5.2. Knowledge pieces extracted from experts' information and probabilities assigned to their nodes in the consequent steps of experts' information processing: Status *"failure".*

#	Knowledge piece maximal node	Nodes	Probability corresponding to the source data	Probability corresponding to the local consistency	Probability corresponding to the internal consistency	Probability corresponding to the external consistency
1	$x_6\ x_7\ x_8$	x_6	[0.2;0.5]	[0.2;0.5]	[0.2;0.5]	[0.2;0.5]
		x_7	[0.6;0.9]	[0.6;0.9]	[0.6;0.9]	[0.6;0.9]
		x_8	[0.5;0.8]	[0.5;0.8]	[0.5;0.8]	[0.5;0.8]
		$x_6\ x_7$	[0.1;0.6]	[0.1;0.5]	[0.1;0.5]	[0.1;0.5]
		$x_6\ x_8$	[0.1;0.5]	[0.1;0.5]	[0.1;0.5]	[0.1;0.5]
		$x_7\ x_8$	[0.4;0.8]	[0.4;0.8]	[0.4;0.8]	[0.4;0.8]
		$x_6\ x_7\ x_8$	[0.0;0.4]	[0.0;0.4]	[0.0;0.4]	[0.0;0.4]
2	$x_6\ x_7\ x_{11}$	x_6	[0.2;0.5]	[0.2;0.5]	[0.2;0.5]	[0.2;0.5]
		x_7	[0.6;0.9]	[0.6;0.9]	[0.6;0.9]	[0.6;0.9]
		x_{11}	[0.6;0.9]	[0.6;0.9]	[0.6;0.9]	[0.6;0.9]
		$x_6\ x_7$	[0.1;0.6]	[0.1;0.5]	[0.1;0.5]	[0.1;0.5]
		$x_6\ x_{11}$	[0.1;0.5]	[0.1;0.5]	[0.1;0.5]	[0.1;0.5]

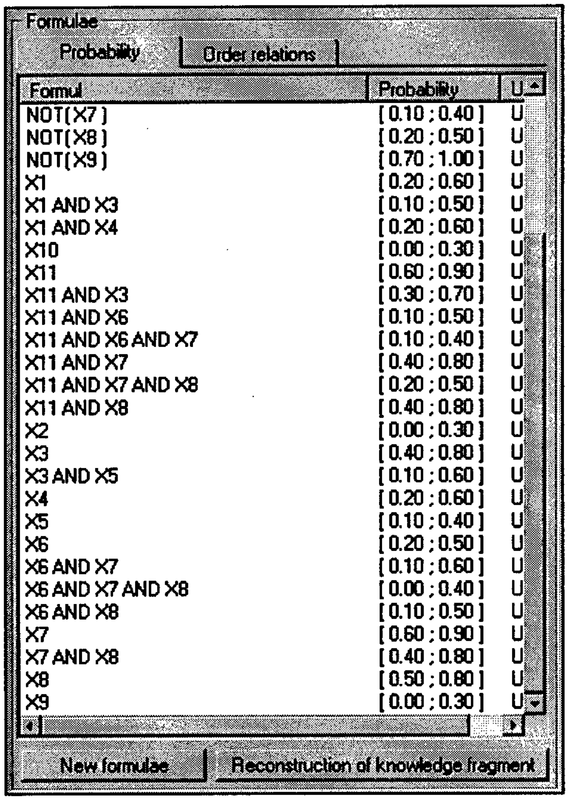
#	Knowledge piece maximal node	Nodes	Probability corresponding to the source data	Probability corresponding to the local consistency	Probability corresponding to the internal consistency	Probability corresponding to the external consistency
		$x_7\ x_{11}$	[0.4;0.8]	[0.4;0.8]	[0.4;0.8]	[0.4;0.8]
		$x_6\ x_7\ x_{11}$	[0.1;0.4]	[0.1;0.4]	[0.1;0.4]	[0.1;0.4]
3	$x_7\ x_8\ x_{11}$	x_7	[0.6;0.9]	[0.6;0.9]	[0.6;0.9]	[0.6;0.9]
		x_8	[0.5;0.8]	[0.5;0.8]	[0.5;0.8]	[0.5;0.8]
		x_{11}	[0.6;0.9]	[0.6;0.9]	[0.6;0.9]	[0.6;0.9]
		$x_7\ x_8$	[0.4;0.8]	[0.4;0.8]	[0.4;0.8]	[0.4;0.8]
		$x_7\ x_{11}$	[0.4;0.8]	[0.4;0.8]	[0.4;0.8]	[0.4;0.8]
		$x_8\ x_{11}$	[0.4;0.8]	[0.4;0.8]	[0.4;0.8]	[0.4;0.8]
		$x_7\ x_8\ x_{11}$	[0.2;0.5]	[0.2;0.5]	[0.2;0.5]	[0.2;0.5]
4	$x_4\ x_{10}\ x_{11}$	x_4	[0,2;0.6]	[0.35;0.6]	[0.35;0.6]	[0.35;0.6]
		x_{10}	[0.0;0.3]	[0.09;0.3]	[0.09;0.3]	[0.09;0.3]
		x_{11}	[0.6;0.9]	[0.6;0.9]	[0.6;0.9]	[0.6;0.9]
		$x_4\ x_{10}$	{0.0, 1.0]	[0.0;0.3]	[0.0;0.3]	[0.0;0.3]
		$x_4\ x_{11}$	{0.0, 1.0]	[0.0;0.6]	[0.0;0.6]	[0.0;0.6]
		$x_{10}\ x_{11}$	{0.0, 1.0]	[0.07;0.3]	[0.07;0.3]	[0.07;0.3]
		$x_4\ x_{10}\ x_{11}$	{0.0, 1.0]	[0.0;0.25]	[0.0;0.25]	[0.0;0.25]
5	$x_1\ x_3$	x_1	[0.2;0.6]	[0.2;0.6]	[0.2;0.6]	[0.2;0.6]
		x_3	[0.4;0.8]	[0.4;0.8]	[0.4;0.8]	[0.4;0.8]
		$x_1\ x_3$	[0.1;0.5]	[0.1;0.5]	[0.1;0.5]	[0.1;0.5]
6	$x_1\ x_4$	x_1	[0.2;0.6]	[0.2;0.6]	[0.2;0.6]	[0.2;0.6]
		x_4	[0.2;0.6]	[0.3;0.6]	[0.35;0.6]	[0.35;0.6]
		$x_1\ x_4$	[0.2;0.6]	[0.2;0.6]	[0.2;0.6]	[0.2;0.6]
7	$x_3\ x_5$	x_3	[0.4;0.8]	[0.4;0.8]	[0.4;0.8]	[0.4;0.8]
		x_5	[0.1;0.4]	[0.1;0.4]	[0.1;0.4]	[0.1;0.4]
		$x_3\ x_5$	[0.1;0.6]	[0.1;0.4]	[0.1;0.4]	[0.1;0.4]
8	$x_3\ x_{11}$	x_3	[0.4;0.8]	[0.4;0.8]	[0.4;0.8]	[0.4;0.8]
		x_{11}	[0.6;0.9]	[0.6;0.9]	[0.6;0.9]	[0.6;0.9]
		$x_3\ x_{11}$	[0.3;0.7]	[0.3;0.7]	[0.3;0.7]	[0.3;0.7]
9	$x_7\ x_9$	x_7	[0.6;0.9]	[0.6;0.9]	[0.6;0.9]	[0.6;0.9]
		x_9	[0.0;0.3]	[0.0;0.3]	[0.0;0.3]	[0.0;0.3]
		$x_7\ x_9$	{0.0, 1.0]	[0.0;0.3]	[0.0;0.3]	[0.0;0.3]
10	x_2	x_2	[0.0;0.3]	[0.0;0.3]	[0.0;0.3]	[0.0;0.3]


Experts' and formal knowledge processing.
Predicates
Formulae
Probability
Order relations
Fragments of knowledge
Add predicate
New formulae
Reconstruction of knowledge fragment
Analysis of a chosen fragment
ABN reconstruction and analysis
Exit

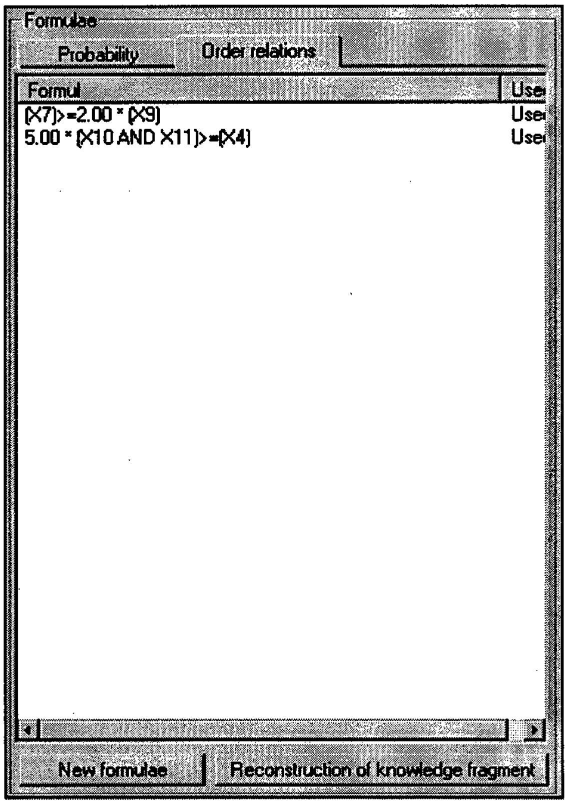
Name	Type	U
X1	User	U
X10	User	U
X11	User	U
X2	User	U
X3	User	U
X4	User	U
X5	User	U
X6	User	U
X7	User	U
X8	User	U
X9	User	U

Formul	Probability	U
NOT(X1)	[0.40 ; 0.80]	U
NOT(X10)	[0.70 ; 1.00]	U
NOT(X11)	[0.10 ; 0.40]	U
NOT(X2)	[0.60 ; 1.00]	U
NOT(X3)	[0.20 ; 0.60]	U
NOT(X4)	[0.30 ; 0.70]	U
NOT(X5)	[0.60 ; 0.90]	U
NOT(X6)	[0.50 ; 0.80]	U
NOT(X7)	[0.10 ; 0.40]	U
NOT(X8)	[0.20 ; 0.50]	U
NOT(X9)	[0.70 ; 1.00]	U
X1	[0.20 ; 0.60]	U
X1 AND X3	[0.10 ; 0.50]	U
X1 AND X4	[0.20 ; 0.60]	U
X10	[0.00 ; 0.30]	U
X11	[0.60 ; 0.90]	U
X11 AND X3	[0.30 ; 0.70]	U
X11 AND X6	[0.10 ; 0.50]	U
X11 AND X6 AND X7	[0.10 ; 0.40]	U
X11 AND X7	[0.40 ; 0.80]	U
X11 AND X7 AND X8	[0.20 ; 0.50]	U
X11 AND X8	[0.40 ; 0.80]	U
X2	[0.00 ; 0.30]	U
X3	[0.40 ; 0.80]	U
X3 AND X5	[0.10 ; 0.60]	U
X4	[0.20 ; 0.60]	U
X5	[0.10 ; 0.40]	U

Know fragments	In Base
X1X3;	
X1X4;	
X10X11X4;	
X11X3;	
X11X6X7;	
X11X7X8;	
X2;	
X3X5;	
X6X7X8;	
X7X9;	



Formul	Probability	U
NOT(X7)	[0.10 ; 0.40]	U
NOT(X8)	[0.20 ; 0.50]	U
NOT(X9)	[0.70 ; 1.00]	U
X1	[0.20 ; 0.60]	U
X1 AND X3	[0.10 ; 0.50]	U
X1 AND X4	[0.20 ; 0.60]	U
X10	[0.00 ; 0.30]	U
X11	[0.60 ; 0.90]	U
X11 AND X3	[0.30 ; 0.70]	U
X11 AND X6	[0.10 ; 0.50]	U
X11 AND X6 AND X7	[0.10 ; 0.40]	U
X11 AND X7	[0.40 ; 0.80]	U
X11 AND X7 AND X8	[0.20 ; 0.50]	U
X11 AND X8	[0.40 ; 0.80]	U
X2	[0.00 ; 0.30]	U
X3	[0.40 ; 0.80]	U
X3 AND X5	[0.10 ; 0.60]	U
X4	[0.20 ; 0.60]	U
X5	[0.10 ; 0.40]	U
X6	[0.20 ; 0.50]	U
X6 AND X7	[0.10 ; 0.60]	U
X6 AND X7 AND X8	[0.00 ; 0.40]	U
X6 AND X8	[0.10 ; 0.50]	U
X7	[0.60 ; 0.90]	U
X7 AND X8	[0.40 ; 0.80]	U
X8	[0.50 ; 0.80]	U
X9	[0.00 ; 0.30]	U



Formul	User
(X7)>=2.00 * (X9)	User
5.00 * (X10 AND X11)>=(X4)	User

Fig.5.8. Printouts of experts' source information and of results of extraction knowledge pieces from experts' information

In tab.5.2 the lines of gray color corresponds to the probabilities which values are changed compared to the information extracted from experts.

5.8. Consistent integration of statistical and expert information within diagnostic model

Algebraic Bayes' Network is a structure for representation of any information under uncertainty. In many practically interesting cases such a way of uncertain knowledge specification is advantageous. Therefore expert information is not only type of information which be represented consistently in the frameworks of ABN. It was mentioned in *Section 4* that knowledge extracted from statistical data contains a lot of sources of uncertainty as well. In fact, frequently we have sampling of statistical data of very small size, in particular, within applications like new hardware diagnostic model design that is the subject of this Project. As a rule probabilistic measure of any statement calculated on the basis of statistical database and/or extracted from expert can be estimated only as confidence interval because point-wise estimations of a probability in the most cases is very inaccuracy and unreliable.

Knowledge extracted from statistical data in a form of logic formulae assigned interval probabilities of truth (they were called classifying predicates in *Section 4*) can be represented in the framework of Algebraic Bayes' Networks like it was described in this section for expert information. Hence, all advantages of ABN-based representation and processing of expert information are valid for knowledge extracted from statistical data. The main advantage of ABN is that statistically estimated confidence intervals combined with experts' assessments of interval probabilities of correlating variables (statements, features, factors) may be narrowed remarkably what corresponds to more high accuracy of a target diagnostic model. A conclusion is that ABN is a convenient framework to join knowledge extracted from experts' information and one assessed over statistical data.

Applied to the new hardware diagnostic model development the above advantage of ABN formal model plays a very significant role. The reason to ascertain its benefit in such application is obvious: insufficiency of statistical data and insufficiency and uncertainty of expert information. To be integrated together in a consistent way within the framework of ABN, they are able to improve significantly diagnostic model.

6. Contribution of the Research and Perspective Future Works

6.1. Contribution of the research

The research presented in this report is aimed at development of mathematical models associated with the technology for information based health assessment system and numerical verification of these models. Up to the time when the real statistical database is accumulated we need mathematical model to generate adequate database that makes it possible to verify developed algorithms and corresponding technology and to do further research aimed at specializing algorithms for new concrete type of device. Of course, each new device will require to develop "ad hoc" model. Nevertheless, the general principles and ideas of development task related model may be borrowed from the Dynamic Data model developed in this research and presented in *Section 2.*

However the major task of this research is development of technology for the accurate assessment of the probability of failure of hardware, such as avionics, on the basis of its known «history of abuse» by environmental and operational factors and assessment of the residual performance resource. The successful solution of both problems allows us to forecast the probability of failure during a forthcoming sortie and to assess the residual performance resource thus providing a quantitative basis for mission planning and timely maintenance as well as preventing emergencies. This application cannot be regarded as a conventional reliability problem because classical reliability does not view exposure to specific environmental conditions and operational factors as a main cause of failures. The problem stated herein does not constitute a conventional prognostic task also because the failure may not occur at all. The problem statement considered in the Report was for the first formulated in the paper [Skormin et al -97]. Such a problem statement is prompted by the modern concept of maintenance known as the «service when needed».

It is expected that the prognostic model presented in this Report is developed on the basis of information downloaded from dedicated monitoring systems of flight-critical hardware and stored in a database. Therefore, the stated problem is related to the area of tasks of Data Mining and KDD ([Frawly et al -91], [Matheus et al 93], [Fayyad et al-95-1], [Fayyad et al 95-2], [Bradley et al-98-1]). According to the existing topics of Data Mining prognostic model design is a classification problem ([Fayyad et al 95-1]). Classification problem of such kind is well known and is being investigated at least during four decades ([Fukunaga-72], [Patrick-72], [Tou et al -74], [Ryin-76]).

Nevertheless, a number of principle tasks of classification are still of great interest and deserve further investigation. For example, a hot area of classification is the so-called problem of feature informativity and algorithms of their selection as well as methods and algorithms of learning for synthesis of classification rule [Bradley et al -98-2]. In addition, there exist a number of problems very important from the applications point of view that still do not have efficient solutions. For example, development of classification models based on Data Mining and KDD for the case when databases contain columns measured on both continuous and discrete scales.

Let us summarize the new results presented in this research and described in [IR-98] and in this Report that constitute its main contribution to the applied classification problem solving and to the area of Data Mining and KDD.

1. The basis of the classification model proposed in the Report is formed by so-called classification predicates. Firstly the idea was proposed by V. Skormin and L. Popyack in [Skormin et al -97]. They are associated with subspaces of factors of low dimension, in particular, with 2-d subspaces. Classification predicates are defined over the entire factor space. Each classification predicate divides the latter into two regions according its truth values (*«true»* and *«false»*) in such way that each region contains mostly realizations of one of two clusters of data. A classification predicate is *true* within a region of the factor space bounded by a set of separation functions of two arguments, which are particular components of the entire factor space. For a 2-d separation bound, the efficient procedure resulting in

optimal separation functions of arbitrary shape (including non- convex case) and associated classification predicates is investigated. This procedure is based on the visualization of cluster projections onto arbitrary 2-d subspaces, and is implemented in an interactive software tool developed in the framework of this research. This procedure provides a user an opportunity to draw any separation rule manually utilizing its approximation by a polygon, i.e. an arbitrary linear spline. Moreover, the regions established by this procedure could be many - connected and non - convex. A user is required to draw a separation bound while the software tool generates the associated classification predicate automatically.

2. A decision tree-like model of the classification procedure is proposed and implemented in numerically efficient software. The peculiarity of decision trees presented in Interim Report [IR-98] in and this Report is that it is binary and consists of a number of ranked classification predicates. Each predicate is associated with a node of the tree and subsets of database realizations that belong to the region of factor space where corresponding classification predicate is *«true»*. The above regions of factor space and subsets of realizations are ranked. The set of regions and corresponding subsets of realizations associated with the leaves of a decision tree are not overlapping and their combination covers the entire factor space and the set of realizations respectively. The last property provides an opportunity to introduce, in a natural way, a set of elementary events and the corresponding probabilistic space that constitutes a model for the assessment of the probability of failure of hardware, i.e. to solve the target task. In addition, the notion of meta – tree used in this research and described in section 4 makes it possible to solve classification task for arbitrary number of clusters of data without any change of the technology of the above decision tree design.
3. The research presented in this paper is aimed at designing a model for reliable assessment of the probability of failure. A pure statistical approach in the case of a small amount of training and testing data is not sufficient for providing the necessary accuracy and reliability of failure prognosis. Therefore, this paper suggests utilizing a number of different decision trees that are supposed to be used to form a more accurate collective decision. Each decision tree consists of a number of ranked classification predicates associated with 2-d subspaces of factors. The most important requirement is that each decision tree has to be associated with different subspaces of factors. Information redundancy is a reason for a possible accuracy enhancement of the assessment of the probability of failure. To employ the idea of redundancy, a special procedure of joint processing of the decisions obtained by individual decision trees is investigated. It is based on the concept of so-called «Algebraic Bayes' Networks» developed by the author ([Gorodetski-92], [Gorodetski et al -97]). In fact, estimation enhancement is achieved due to utilizing the background knowledge. This method is demonstrated numerically. Additional utilization of interval mathematics methods to calculate the posterior probability of failure on the basis of Bayes' formula makes it possible to obtain the precise upper (minimal) bound of the target probability.

   It should be noted that the classification rule development technique presented could be efficiently used in a wide area of applied tasks of Data Mining and KDD.
4. The utilization of the concept of a classification predicate defined over subspaces of low dimensions makes possible to develop a totally new approach to the task of rules extraction from databases in the most complex case. One such case is the situation when a database contains columns specified both in continuous and discrete scales. It is well known that now this task is one of the key problems of Data Mining and KDD. Any known and conventionally used approach aiming at the creation of a knowledge base by «mining» a database containing both continuous and discrete data is based on direct discretization of continuous (real valued) data and results in substantial dimension increase of the factor space. Consequently, such an approach leads to inefficient algorithms and can not be recognized as a satisfactory one even if a discretiation is made optimally.

As an alternative, an approach based on the utilization of the concept of classification predicate makes it possible to avoid artificial discretization at all. Actually, a classification predicate itself defined over a subset of continuous factors (features) can be considered as a discrete specification of continuous data. A classification predicate can be considered as a new feature which represents the same data in a new way. There exist a number of known approaches to cope with the task of extraction rules from database with columns specified in discrete scales [see, for example, [Michalski et al -81], [Quinlian-83], [Michalski-90]). Hence, the concept of classification predicate makes it possible to solve a number of difficult Data Mining and KDD problems in an efficient new way.

One more original approach to solve the task of rule extraction from learning data was proposed by author of this Report [Gorodetski et al -96]. It was described in brief in Appendix of the Interim Report [IR-98].

5. Algebraic Bayes' Networks theory developed by author and presented in Interim Report [IR-98] and in this one possess a number of advantages regarding to the application that is the subject of this research and regarding to a more wide area of Knowledge Engineering. The area of its applications is dealing with uncertain and sub-defined data including expert's knowledge.

6.2. Proposals for future research

This research may be considered as a step in the development of technology for information-based health assessment system design. Of course it is not able to solve a number of problems associated with this very important and difficult task. However, there may be pointed out a number of theoretical and applied problems that in my opinion has to be a subject of research in the framework of health assessment system design. They are as follows:

1. *Mathematical model for mining knowledge from database of multi-scale data structures.* In compare with the model developed in this research, the proposed research aims at development a mathematical model of technology which integrates (1) processing real valued database resulting in obtaining classification predicates and (2) processing discrete valued data resulting in extraction rules from both real valued and discrete valued data. This research may aim at development a mathematical basis for advanced data mining technology applicable for wide area of information-based health assessment systems.
2. *Development of software tool prototype aimed at supporting an interactive technology of information-based health assessment system.* Development of such tool could make it possible to investigate numerically the pros and cons of any mathematical basis, its advantages and deficiencies. This tool may be used to prepare future developers to implement technology. It might be a first step to development of powerful multi-purpose software tool for utilization in the area of information –based health assessment systems design.
3. *Advanced statistical and logical models for extracting sensitive patterns from database.* This mathematical model aims at solving such practically important prognostic related tasks as:

- ranking particular environmental conditions as factors responsible for general and particular types of failures,
- determination of particular groups of environmental conditions ("patterns") and assessment of their combined effects on failures in general and on particular types of failures,
- justification of the development of devices protecting from adverse environmental conditions,
- development of the recommendations on the avoidance of the combined effects of adverse conditions.

Conventionally these tasks are solved by methods of mathematical statistics which uses ideas from component and factor analyses. But the latter are not appropriate in many practical situations. In addition, in these tasks it is necessary to deal with continuous and discrete factors what takes to develop a more powerful mathematically justified approach. It should be noted that this problem now is the subject of intensive research in Data Mining area.

References

[Anderson-60]. T.W.Anderson. (1960). An Introduction to Multivariate Statistical Analysis. Chapmen & Hall, Ltd.

[Birkhoff-67]. G.Birkhoff. (1967), Lattice Theory. Providence, Rhode Island, 1967.

[Bradley et al-98-1]. P.Bradley, U.Fayyad, O.Mangasarian. (1998), Data Mining: Overview and Optimization Opportunity. *http://www.research.microsoft.com/datamine/.*

[Dempster-66]. A.P. Dempster. (1966), Upper and Lower Probabilities induced by a Multi-valued Mapping. *Annals of Mathematical Statistics*, 36, 1966, pp.325-339.

[Dubois-Prade-88]. D.Dubois, H.Prade.(1988), The'orie des Possibilities.Applications a' la Representationdes Connaissances en Informatique. MASSON, Paris.

[Fagin et al-88]. J.Y.Halpern, N.A.Megiddo. (1988), Logic for Reasoning about Probabilities. In *Proceedings 3th IEEE Symposium on Logic and Computer Science*, 1988. (The extended version of the paper is published as Report RJ6190(60900), 4/12/, pp.1-41).

[Fagin et al-89]. R.Fagin, J.Y.Halpern. (1989), Uncertainty, Belief, and Probability. In *Proceedings of 11th International Joint Conference on Artificial Intelligence.* Morgan Kaufman Publ. Detroit, Michigan, USA, pp.1161-1167. [Dubois-Prade-88]. D.Dubois, H.Prade.(1988), The'orie des Possibilities.Applications a' la Representationdes Connaissances en Informatique. MASSON, Paris.

[Fayyad et al-95-1]. U.M.Fayyad, G.Piatetsky-Shapiro, P.Smyth. (1995), From Data Mining to Knowledge Discovery: An Overview. In *"Advances in Knowledge Discovery and Data Mining"* (Eds. U.M.Fayyad, G.Piatetsky-Shapiro, P.Smyth), Cambridge, Mass: MIT Press, pp. 1-34.

[Fayyad et al 95-2]. U.M.Fayyad, G.Piatetsky-Shapiro, P.Smyth. (1995), From Data Mining to Knowledge Discovery: An Overview. In *"Advances in Knowledge Discovery and Data Mining"* (Eds. U.M.Fayyad, G.Piatetsky-Shapiro, P.Smyth), Cambridge, Mass: MIT Press, pp. 1-34.

[Frawley et al-91]. W.J.Frawley, G.Piatetsky-Shapiro, and C.J.Matheus.(1991), Knowledge Discovery in Data Bases: An Overview. In *"Knowledge Discovery in Data Bases"* (Eds. G.Piatetsky-Shapiro and W.J.Frawley), Cambridge, Mass:AAAI/MIT Press, pp.1-27.

[Fukunaga-72]. K.Fukunaga. (1972), Introduction to Statistical Pattern Recognition. Academic Press, New York.

[Gorodetski et al-96]. V.Gorodetski, O.Karsaev. (1996), Algorithm of Rules Extraction from Learning Data. *In Expert Systems Applications &Artificial Intelligence (EXPERSYS-96), Technology Transfer Series (ed. A.Niku-Lary),* France, pp.133-138.

[Gorodetski-92]. V.Gorodetski. (1992), Adaptation Problems in Expert Systems. *In International Journal of Adaptive Control and Signal Processing.* Vol.6, pp.201-209.

[Gorodetski et al-97]. V.Gorodetski, A.Tulupiev. (1997), Development a Consistent Knowledge Base under Uncertainty. *Transactions. of the Russian Academy of Sciences "Theory and Control Systems"*, Vol.5, pp. 33-42.

[IR-98] V.Gorodetski. (1998). Applied Methods and Models of Knowledge Engineering in Information Based Health Assessment Systems. Interim Report, Contract No. F61775-98-WE116), St. Petersburg, Russia, 80 pp.

[Matheus et al 93]. C.J.Matheus, P.Chan, and G.Piatetsky-Shapiro. (1993), Systems for Knowledge Discovery. *IEEE Trans. On Knowledge and Data Engineering*, 5 (6), pp.903-913.

[Michalski et al-81]. R.S.Michalski, T.C.Dietterich. (1981), Inductive Learning of Structural Descriptions: Evaluation Criteria and Comparative Review of Selected Method. *Artificial Intelligence,* v.16, 3.

[Michalski-90]. R.S.Michalsky. (1990), Learning Flexible Concepts: Fundamental Ideas and Methodology. In *Machine Learning: An Artificial Intelligence Approach.* v.III. (Eds. Y.Kondratoff and R.S.Michalsky), Morgan Kaufmann Publishers.

[Patrick-72]. E.Patrick (1972), Fundamentals of Pattern Recognition. Prentice-Hall, Inc. Englewood Cliffs, N.J.

[Popyack-98]. L.Popyack. (1998), Ph.D.Thesis. Binghamton University, Binghamton, NY, USA.

[Popyack-99]. L.J. Popyack, V.A. Skormin, V.I. Gorodetski, M.L. Araiza, J.D. Michel, «Applications of Cluster Analysis in Diagnostics-Related Problems,» Proceedings of the 1999 IEEE Aerospace Conference, Snowmass at Aspen, Colorado, 6-13 March 1999.

[Quinlian-83]. J.R.Quinlian. (1983), Inductive Inference as a Tool for the Construction of High-Performance Programs. In *Machine Learning: An Artificial Intelligence Approach.* (ed. R.S. Michalsky, J.G.Carbonell, T.M.Mitchell),-Palo Alto, Tioga Publishing Company.

[Rao-67]. C.R.Rao. (1968) Linear Statistical Inference and its Applications. John Willey &Sons, Inc.

References

[Ryin-76]. Van Ryin (Ed). (1976),. *Proceedings of Advanced Seminar, Conducted by the Mathematics Research Center*. The University of Wisconsin at Madison.

[Shafer-76]. G.Shafer. (1976), A Mathematical Theory of Evidence. Princeton University Press.

[Skormin et al -97]. V.Skormin, L.Popyack. (1997), Reliability of Avionics and "History of Abuse". A Prognostic Technique. In *Proceedings of ICI&C '97,* St. Petersburg, Russia. pp. Lxxvi-lxxxii.

[Tou et al-74]. J.T.Tou, R.C. Gonzales. (1974), Pattern Recognition Principles. Addison-Wesley Publishing Company.

[Zadeh-78]. L.A.Zadeh. (1978), Fuzzy Sets as a Basis for a Theory of Possibility. In *Fuzzy Sets Systems*, v.1, No.4, pp.395-460.

[WWW-1]. Web site of Data Sets. *http://www.kdnuggets.com/datasets.html.*

[WWW-2]. Learning Data Web site. *http://www.ics.uci.edu/AI/ML/Machine-Learning.html.*

Appendix A1. Trajectories of Failure Development

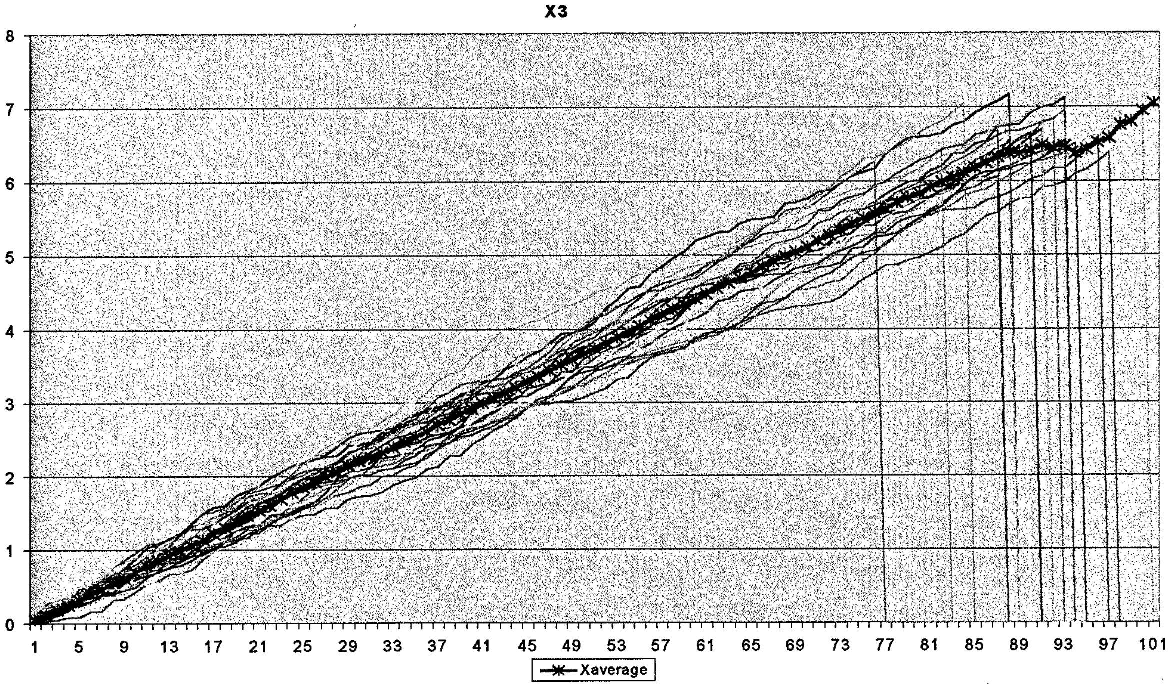


Fig.A1.1. Realization of trajectories of development of adverse exposure X_3 as the functions of the number of aircraft sortie

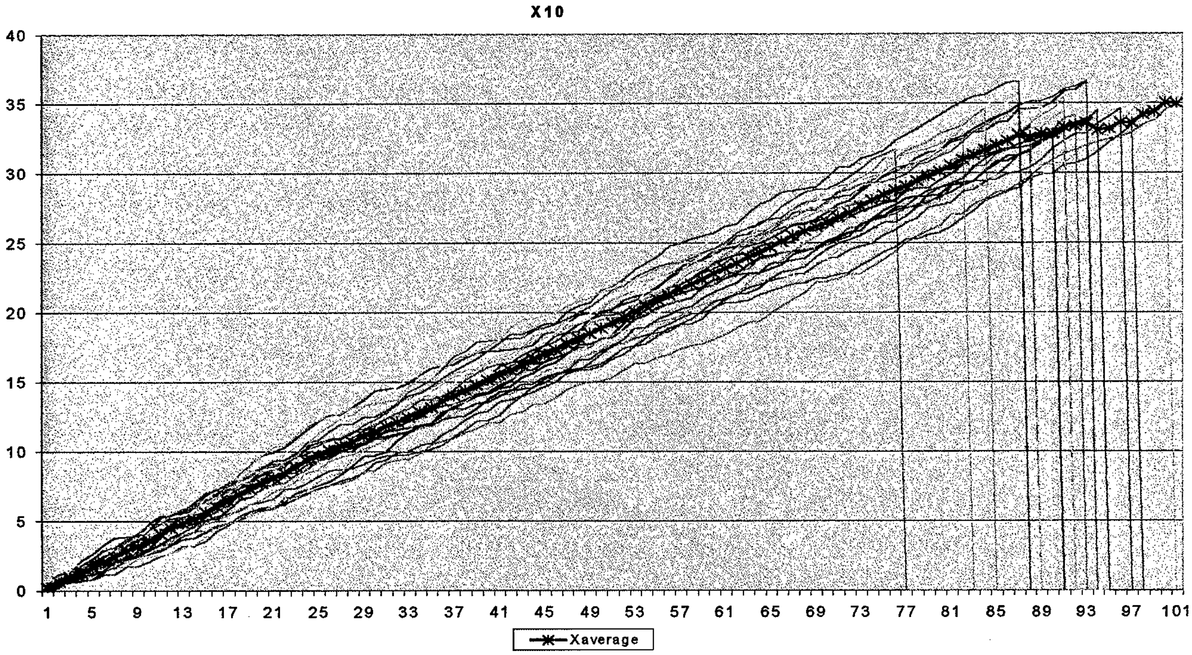


Fig.A1.2. Realization of trajectories of development of adverse exposure X_{10} as the functions of the number of aircraft sortie

Appendix A1. Trajectories of failure development

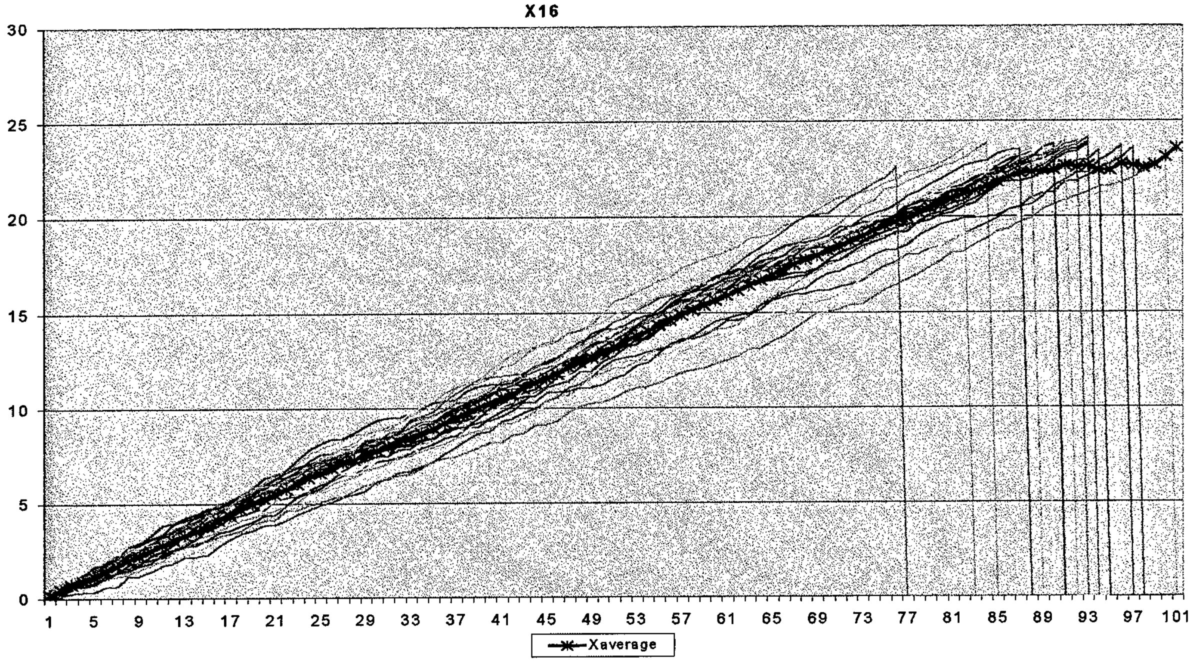


Fig.A1.3. Realization of trajectories of development of adverse exposure X_{16} as the functions of the number of aircraft sortie

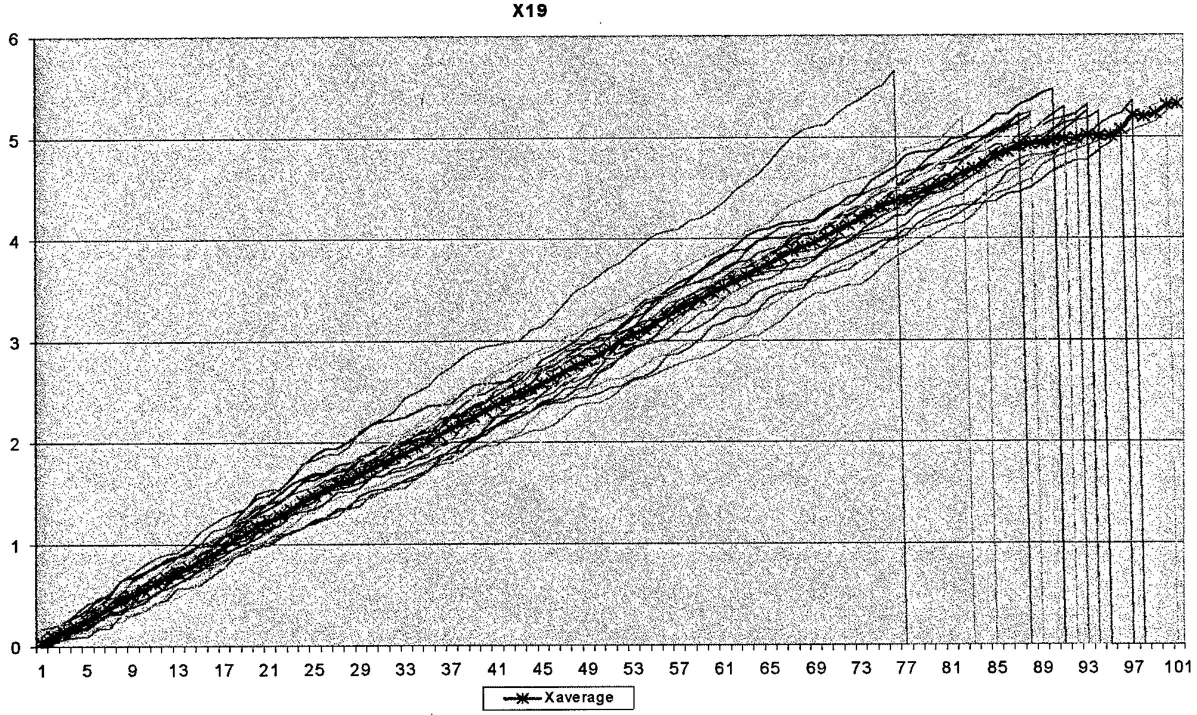


Fig.A1.4. Realization of trajectories of development of adverse exposure X_{19} as the functions of the number of aircraft sortie

Appendix A1. Trajectories of failure development

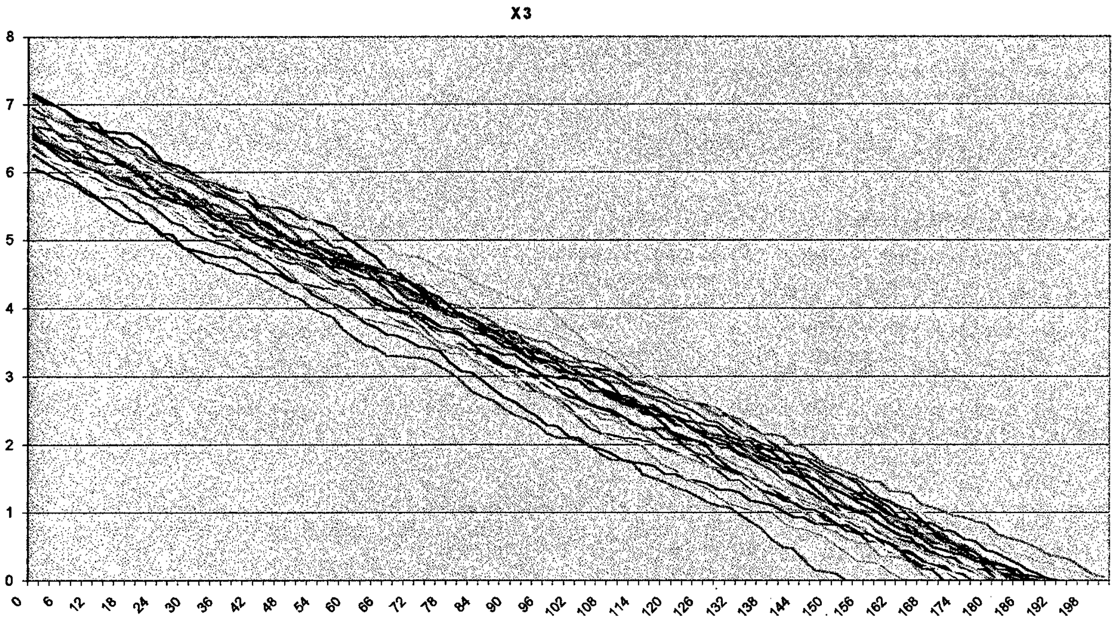


Fig.A1.5. Realization of trajectories of development of adverse exposure X_3 as the functions of residual performance resource

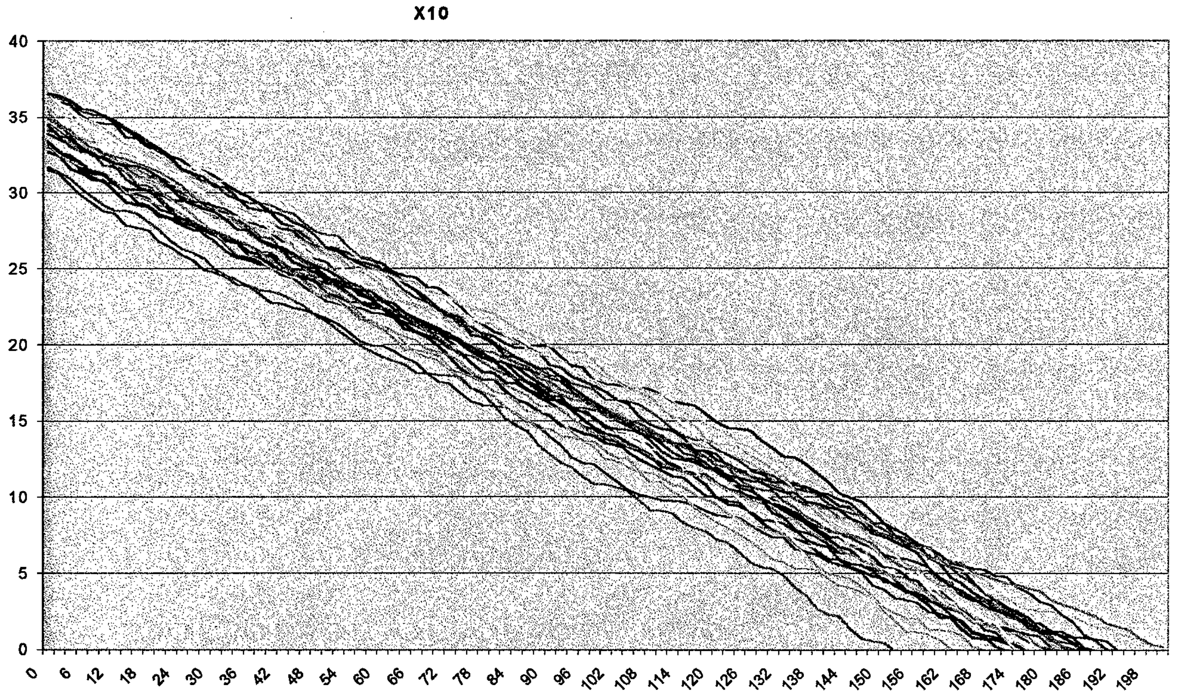


Fig.A1.6. Realization of trajectories of development of adverse exposure X_{10} as the functions of residual performance resource

Appendix A1. Trajectories of failure development

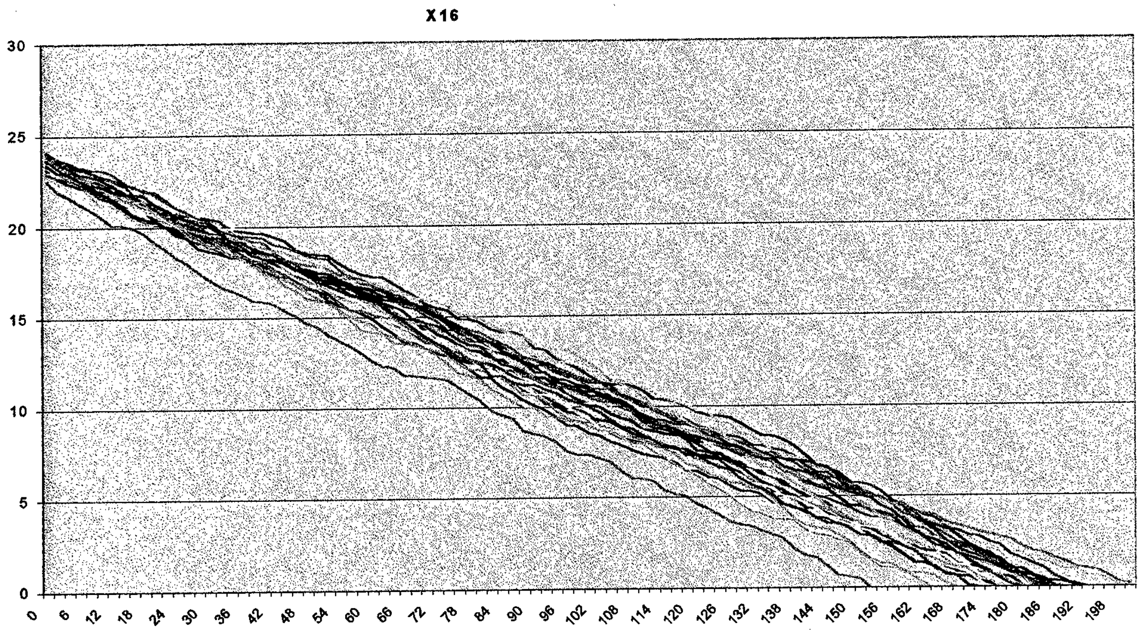


Fig.A1.7. Realization of trajectories of development of adverse exposure X_{16} as the functions of residual performance resource

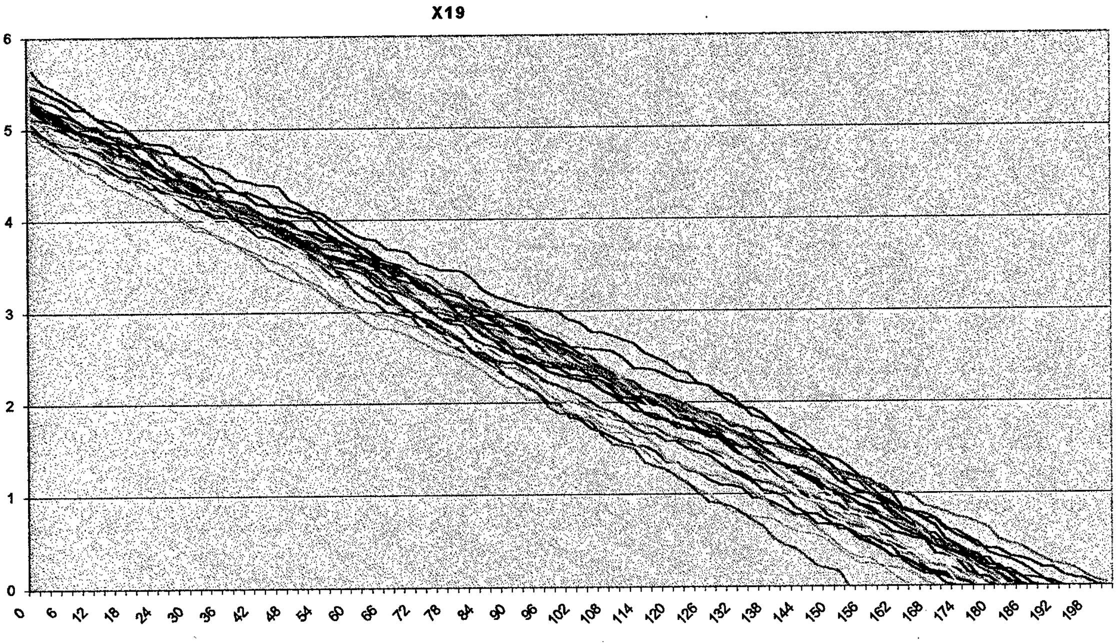


Fig.A1.8. Realization of trajectories of development of adverse exposure X_{19} as the functions of residual performance resource

Appendix A2. Learning and Testing Data

Learning data

Nr	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19	X20	Stat
1	18,399	4,494	6,647	33,033	12,823	40,052	11,671	10,360	10,662	32,749	29,208	12,006	41,504	29,032	26,723	23,842	16,046	4,700	5,476	0	1
2	6,812	1,298	2,118	10,799	4,307	13,249	3,992	3,460	3,466	10,542	10,006	4,067	14,572	9,614	8,704	7,912	5,180	1,475	1,819	122	-1
3	13,832	3,293	4,824	24,140	9,438	28,951	8,981	7,720	7,938	24,056	22,532	8,814	31,570	21,976	19,504	17,506	11,703	3,344	4,100	50	-1
4	10,621	2,249	3,350	17,561	6,785	20,451	6,490	5,740	5,792	17,528	15,798	6,411	22,912	15,873	14,344	12,669	8,468	2,374	2,965	82	-1
5	6,617	1,242	2,023	10,095	4,234	12,688	3,845	3,341	3,342	10,093	9,579	3,930	13,667	9,382	8,426	7,642	5,011	1,385	1,772	126	-1
6	17,495	4,271	6,230	31,043	12,154	37,692	11,136	9,783	10,101	30,898	27,983	11,322	38,923	27,776	25,338	22,547	15,169	4,370	5,231	8	0
7	7,364	1,388	2,246	11,507	4,647	14,042	4,264	3,745	3,763	11,548	10,806	4,281	15,654	10,607	9,254	8,438	5,548	1,597	1,972	118	-1
8	18,086	4,369	6,457	32,287	12,474	38,880	11,532	10,146	10,380	31,645	28,584	11,769	40,699	28,447	26,025	23,098	15,545	4,537	5,371	4	0
9	17,693	4,107	5,692	29,988	12,193	42,340	9,440	8,664	9,363	32,701	31,572	10,117	43,233	25,684	23,787	21,678	14,848	4,674	4,849	10	0
10	3,871	0,953	1,114	5,611	2,835	9,439	1,941	1,736	1,871	7,418	5,958	2,038	9,573	5,629	4,907	4,446	3,124	1,081	0,911	138	-1
11	5,423	1,211	1,475	8,168	3,607	12,848	2,672	2,325	2,472	9,449	8,831	2,824	12,951	7,480	6,784	6,028	4,134	1,442	1,242	124	-1
12	1,909	0,444	0,485	2,480	1,363	4,267	1,016	0,851	0,858	3,325	2,512	0,975	5,173	2,226	2,231	1,940	1,364	0,528	0,402	162	-1
13	11,290	2,501	3,424	19,265	7,311	27,330	5,619	4,899	5,415	20,075	19,194	6,130	28,480	15,691	13,870	12,877	8,712	3,034	2,840	74	-1
14	19,101	4,391	6,063	31,685	13,210	45,114	10,301	9,325	10,084	34,894	33,779	10,892	46,629	27,567	25,714	23,216	15,938	4,984	5,235	0	1
15	17,988	4,167	5,812	30,445	12,457	43,184	9,642	8,812	9,567	33,200	32,283	10,334	44,353	26,202	24,169	22,082	15,147	4,761	4,940	8	0
16	6,488	1,507	1,736	9,878	4,288	14,815	3,284	2,885	3,104	11,283	10,593	3,392	16,226	8,506	7,908	7,206	4,950	1,671	1,529	114	-1
17	5,794	1,483	2,333	11,013	4,463	16,132	3,630	3,024	3,285	12,932	11,181	3,451	17,019	8,995	8,802	7,913	5,476	1,961	1,562	122	-1
18	17,511	4,645	6,639	30,171	13,646	45,577	10,971	9,195	9,705	34,632	31,742	11,047	45,169	26,895	26,142	23,443	15,746	5,145	4,912	0	1
19	7,313	1,910	2,899	14,121	5,523	20,449	4,675	3,771	4,064	15,895	13,777	4,457	21,827	11,292	10,751	9,819	6,774	2,486	1,931	104	-1
20	10,691	2,645	4,043	19,238	7,928	28,214	6,672	5,394	5,732	21,753	19,255	6,427	30,372	15,686	15,493	13,694	9,264	3,317	2,831	72	-1
21	17,471	4,629	6,600	30,052	13,595	45,451	10,904	9,140	9,654	34,517	31,568	10,994	45,028	26,790	26,009	23,333	15,675	5,134	4,886	2	0
22	2,986	0,746	1,239	5,582	2,417	8,586	1,902	1,571	1,767	7,138	6,133	1,759	9,056	4,332	4,612	4,482	3,058	1,059	0,802	156	-1
23	17,143	4,551	6,400	29,425	13,250	44,431	10,651	8,876	9,415	33,798	31,012	10,681	44,084	26,172	25,305	22,761	15,271	5,014	4,763	6	0
24	0,629	0,168	0,269	1,078	0,547	2,075	0,483	0,285	0,348	1,327	1,490	0,418	1,895	0,951	1,000	0,972	0,624	0,214	0,163	180	-1
25	5,387	1,447	2,080	8,327	4,454	13,138	3,724	3,172	3,346	9,715	11,679	3,735	13,658	7,957	8,495	7,917	5,273	1,284	1,759	114	-1
26	15,931	4,193	6,146	26,018	12,841	38,944	10,453	9,142	9,620	29,358	30,783	10,804	41,393	24,261	23,963	22,128	14,853	3,955	5,016	8	0
27	16,402	4,314	6,351	26,973	13,226	40,205	10,675	9,398	9,906	30,542	31,639	11,025	42,861	25,478	24,528	22,655	15,269	4,112	5,148	4	0

Appendix A2. Learning and testing data

N	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19	X20	Stat
28	16,896	4,479	6,515	27,735	13,595	41,352	10,930	9,579	10,102	31,604	32,323	11,196	44,568	26,484	24,959	22,963	15,470	4,276	5,254	0	1
29	5,481	1,465	2,116	8,447	4,531	13,319	3,785	3,237	3,413	9,849	11,951	3,800	13,881	8,077	8,649	8,065	5,363	1,293	1,794	110	-1
30	13,173	3,523	5,198	22,077	10,645	33,003	8,741	7,676	8,118	24,509	26,246	9,171	34,674	19,878	20,342	18,857	12,620	3,291	4,233	32	-1
31	4,959	1,273	1,913	7,498	4,090	12,093	3,403	2,896	3,071	8,870	10,584	3,449	12,369	7,287	7,883	7,278	4,836	1,177	1,612	120	-1
32	2,487	0,632	0,904	3,624	2,014	5,584	1,763	1,559	1,588	4,242	5,116	1,781	6,196	3,565	4,090	3,564	2,417	0,540	0,817	148	-1
33	15,878	4,104	6,314	28,635	12,551	40,552	10,515	8,963	9,651	31,287	30,008	10,601	40,549	26,986	24,132	22,139	14,983	4,516	4,947	8	0
34	8,578	2,093	3,192	15,524	6,203	19,511	5,519	4,797	5,235	15,905	15,303	5,535	19,913	14,573	12,663	11,541	7,744	2,178	2,692	94	-1
35	10,361	2,619	3,993	18,823	7,802	24,906	6,871	5,902	6,386	19,624	18,786	7,006	25,055	17,516	15,927	14,325	9,624	2,766	3,295	76	-1
36	9,952	2,478	3,734	17,990	7,338	23,348	6,500	5,599	6,095	18,788	17,727	6,570	23,695	16,618	15,143	13,594	9,147	2,612	3,138	82	-1
37	16,947	4,367	6,704	30,431	13,340	43,094	11,203	9,594	10,270	33,066	31,896	11,379	42,886	29,085	25,605	23,508	15,978	4,758	5,302	0	1
38	1,587	0,484	0,654	3,201	1,163	3,623	1,126	1,006	1,067	2,910	3,388	1,124	4,123	2,858	2,481	2,182	1,416	0,339	0,590	160	-1
39	16,364	4,197	6,359	29,087	12,769	41,275	10,681	9,101	9,777	31,724	30,452	10,796	41,563	27,459	24,491	22,383	15,140	4,583	5,034	6	0
40	0,056	0,021	0,025	0,102	0,055	0,174	0,038	0,031	0,040	0,154	0,170	0,030	0,185	0,093	0,114	0,090	0,063	0,018	0,020	180	-1
41	17,092	4,852	7,004	34,481	12,841	44,399	12,120	9,256	10,086	32,466	31,410	11,952	47,807	29,579	23,820	22,985	15,379	5,060	5,090	4	0
42	17,324	4,892	7,077	34,749	13,012	44,896	12,283	9,353	10,183	32,726	31,575	12,126	48,500	29,799	24,020	23,190	15,514	5,107	5,148	2	0
43	13,161	3,666	5,439	26,248	9,839	33,569	9,442	7,224	7,775	24,777	24,182	9,202	36,761	22,200	18,235	17,678	11,654	3,773	3,978	48	-1
44	17,577	4,973	7,169	35,147	13,241	45,666	12,415	9,454	10,291	33,420	31,951	12,226	49,490	30,098	24,415	23,424	15,693	5,195	5,210	0	1
45	7,687	2,046	2,980	14,464	5,520	18,386	5,515	4,000	4,144	13,576	13,501	5,086	21,467	12,547	9,906	9,392	6,151	2,053	2,274	100	-1
46	6,776	1,733	2,601	12,942	4,762	15,896	4,814	3,524	3,703	12,204	11,798	4,359	18,520	11,248	8,794	8,267	5,466	1,814	1,994	112	-1
47	1,873	0,378	0,567	3,486	1,065	3,793	1,161	0,790	0,814	2,947	2,220	1,029	4,691	2,795	1,918	1,825	1,166	0,494	0,410	158	-1
48	0,665	0,120	0,215	1,281	0,358	1,403	0,524	0,258	0,264	0,711	0,838	0,459	1,510	1,005	0,635	0,625	0,383	0,162	0,145	170	-1
49	11,989	3,060	4,288	20,866	8,807	26,284	7,841	6,620	7,090	22,820	18,981	7,432	28,736	18,279	17,434	16,158	10,881	3,249	3,456	68	-1
50	5,336	1,331	1,765	8,758	3,871	11,230	3,357	2,843	3,100	9,915	8,412	3,147	13,207	7,706	7,396	6,734	4,586	1,397	1,497	140	-1
51	17,514	4,522	6,550	30,598	13,183	39,350	11,838	9,851	10,552	33,913	30,252	11,015	44,980	26,830	25,152	23,914	15,927	4,658	5,319	0	1
52	17,199	4,393	6,401	29,984	12,902	38,519	11,558	9,653	10,327	33,279	29,479	10,760	43,907	26,184	24,734	23,449	15,626	4,589	5,195	4	0
53	5,168	1,275	1,677	8,421	3,716	10,858	3,138	2,728	2,960	9,460	8,069	3,022	12,836	7,276	7,132	6,382	4,366	1,343	1,431	142	-1
54	1,161	0,309	0,336	1,649	0,859	2,615	0,845	0,544	0,516	1,836	1,415	0,736	2,533	1,373	1,762	1,356	0,955	0,326	0,292	176	-1
55	16,444	4,211	6,061	28,113	12,413	36,841	11,010	9,161	9,767	31,690	28,161	10,203	41,644	24,807	23,635	22,350	14,859	4,339	4,937	14	-1
56	16,931	4,311	6,276	29,542	12,651	37,740	11,336	9,517	10,168	32,628	28,904	10,597	43,060	25,678	24,316	23,044	15,357	4,499	5,117	8	0
57	13,468	3,539	5,251	24,913	10,332	33,980	8,377	7,397	8,046	27,791	23,738	8,484	34,847	21,246	20,168	18,131	12,293	4,011	3,993	44	-1

Appendix A2. Learning and testing data

N	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19	X20	Stat
58	8,695	2,229	3,120	14,198	6,674	19,860	5,428	4,733	5,062	17,234	14,845	5,151	20,893	12,969	13,139	11,255	7,548	2,328	2,583	106	-1
59	1,867	0,438	0,636	2,729	1,399	4,250	1,263	1,011	1,081	3,311	3,254	1,207	4,299	2,680	2,900	2,396	1,599	0,499	0,511	170	-1
60	12,782	3,371	4,942	23,690	9,732	31,956	7,991	6,978	7,589	26,284	22,455	7,991	32,725	20,244	19,163	17,087	11,559	3,799	3,771	52	-1
61	15,368	4,034	6,144	28,624	11,905	39,353	9,519	8,411	9,122	32,218	27,572	9,561	40,718	24,024	23,142	20,713	13,956	4,666	4,550	24	-1
62	16,449	4,406	6,756	31,398	12,898	43,313	10,253	9,072	9,851	34,636	30,308	10,514	43,845	26,280	25,216	22,453	15,140	5,052	4,960	10	0
63	17,450	4,642	7,137	33,004	13,706	45,636	10,851	9,578	10,441	36,596	31,867	11,155	46,059	27,473	26,830	23,850	16,027	5,286	5,269	0	1
64	16,768	4,467	6,921	32,012	13,154	43,961	10,582	9,274	10,057	35,150	30,665	10,817	44,555	26,687	25,822	22,934	15,450	5,129	5,076	6	0
65	1,612	0,441	0,545	2,592	1,177	3,284	1,063	0,911	0,880	2,572	2,020	1,078	4,284	2,427	2,183	1,757	1,179	0,412	0,424	172	-1
66	7,370	2,089	2,738	14,102	5,410	17,747	4,669	4,090	4,300	13,758	13,042	4,849	19,623	11,146	10,991	9,452	6,469	1,994	2,203	110	-1
67	18,275	4,728	6,161	31,510	12,945	42,754	11,217	9,439	9,881	32,244	31,400	11,486	46,165	25,958	25,533	22,796	15,295	4,779	5,196	4	0
68	18,419	4,805	6,265	31,920	13,142	43,500	11,319	9,508	9,988	32,792	31,723	11,608	47,041	26,195	25,713	23,093	15,495	4,863	5,243	2	0
69	18,748	4,878	6,395	32,453	13,376	44,242	11,563	9,729	10,200	33,458	32,668	11,804	47,629	26,825	26,329	23,632	15,823	4,949	5,362	0	1
70	2,072	0,583	0,693	3,537	1,460	4,042	1,415	1,192	1,141	3,276	2,875	1,356	5,653	2,971	2,789	2,204	1,524	0,477	0,576	166	-1
71	2,910	0,973	1,188	5,940	2,225	6,998	2,017	1,728	1,728	5,739	5,592	1,889	8,793	4,791	4,320	3,611	2,414	0,831	0,913	156	-1
72	5,199	1,443	2,003	9,870	3,863	12,055	3,320	2,965	3,080	9,946	8,966	3,377	13,760	7,783	7,965	6,634	4,495	1,407	1,596	130	-1
73	2,856	0,608	1,038	5,364	2,007	7,632	1,867	1,461	1,544	5,385	4,989	1,878	6,953	4,940	4,295	3,870	2,771	0,895	0,761	134	-1
74	15,018	3,492	5,251	25,269	11,008	36,466	9,711	7,882	8,213	27,364	27,023	9,472	37,234	22,556	21,798	19,675	13,602	4,009	4,343	30	-1
75	7,459	1,554	2,641	12,881	5,327	18,120	4,801	3,832	4,075	13,381	13,741	4,639	18,381	11,332	10,799	9,708	6,756	1,981	2,093	98	-1
76	13,645	3,203	5,042	23,378	10,326	34,016	9,008	7,418	7,756	25,597	25,543	8,861	33,825	21,227	20,468	18,878	13,023	3,748	4,080	38	-1
77	16,783	3,986	6,058	29,003	12,469	41,782	10,649	8,847	9,372	31,584	30,816	10,598	42,105	25,923	24,500	22,441	15,419	4,579	4,905	8	0
78	16,939	4,019	6,110	29,257	12,569	42,069	10,773	8,915	9,449	31,802	31,055	10,700	42,471	26,077	24,728	22,608	15,526	4,622	4,946	6	0
79	17,472	4,200	6,448	30,511	13,098	43,287	11,427	9,411	9,956	33,169	32,584	11,162	44,361	27,134	25,916	23,641	16,253	4,770	5,206	0	1
80	8,328	1,818	3,118	14,703	6,201	20,752	5,522	4,500	4,788	15,586	15,528	5,417	21,448	12,956	12,460	11,235	7,870	2,286	2,452	88	-1
81	8,383	2,395	3,203	16,038	6,123	20,516	5,291	4,861	5,024	17,667	14,167	5,214	20,960	13,842	12,905	11,383	7,833	2,649	2,458	98	-1
82	17,172	4,681	6,280	32,235	12,130	40,410	10,793	9,620	9,793	34,419	29,310	10,316	42,508	27,189	25,505	22,254	15,185	5,275	4,855	10	0
83	15,970	4,403	5,895	29,770	11,415	37,391	10,200	9,084	9,244	32,494	27,393	9,610	39,328	25,586	23,794	21,005	14,333	4,866	4,587	20	-1
84	14,480	3,967	5,457	27,649	10,410	35,242	9,092	8,172	8,448	30,350	24,682	8,778	35,814	23,604	21,870	19,354	13,306	4,576	4,129	36	-1
85	14,330	3,918	5,369	27,301	10,238	34,780	8,944	8,002	8,275	29,871	24,383	8,617	35,282	23,285	21,503	18,992	12,999	4,519	4,057	38	-1
86	17,820	4,791	6,461	33,405	12,500	41,826	11,119	9,874	10,052	35,440	30,437	10,611	43,838	28,101	26,331	22,871	15,617	5,443	4,993	4	0
87	18,242	4,918	6,743	34,980	12,789	43,518	11,464	10,176	10,375	36,622	31,193	11,031	45,155	28,919	27,194	23,766	16,272	5,645	5,114	0	1

Appendix A2. Learning and testing data

N	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19	X20	Stat
88	4,213	1,108	1,346	7,241	2,825	9,408	2,396	2,179	2,156	8,186	6,514	2,262	9,657	6,446	5,776	4,947	3,375	1,272	1,063	142	-1
89	17,170	4,674	6,234	31,869	12,699	42,089	9,900	9,397	9,642	35,426	31,010	9,883	45,500	24,824	24,988	22,928	15,583	5,065	4,830	4	0
90	6,179	1,606	2,249	12,464	4,359	16,235	3,311	3,111	3,299	12,779	11,072	3,553	15,656	9,387	8,692	8,278	5,658	1,840	1,644	112	-1
91	4,462	1,215	1,709	9,231	3,202	11,371	2,535	2,336	2,506	9,169	8,218	2,651	11,602	6,967	6,418	5,993	4,096	1,248	1,279	126	-1
92	1,170	0,374	0,364	2,293	0,817	2,740	0,596	0,648	0,630	2,513	2,217	0,581	3,109	1,814	1,687	1,549	1,120	0,337	0,310	164	-1
93	0,272	0,120	0,092	0,614	0,225	1,064	0,121	0,132	0,115	0,651	0,694	0,149	1,246	0,308	0,336	0,364	0,282	0,126	0,052	172	-1
94	3,761	1,136	1,510	7,812	2,846	9,587	2,417	2,139	2,247	7,854	7,087	2,389	10,335	6,108	5,600	5,236	3,641	1,068	1,147	132	-1
95	16,856	4,595	6,142	31,512	12,451	41,476	9,770	9,261	9,491	34,744	30,560	9,791	44,577	24,398	24,641	22,661	15,404	4,991	4,741	6	0
96	17,774	4,854	6,512	32,830	13,288	43,504	10,381	9,854	10,129	36,951	32,121	10,327	47,576	25,904	26,007	23,936	16,311	5,269	5,049	0	1
97	15,901	4,502	6,967	28,897	13,586	41,333	11,471	9,683	10,439	35,704	30,973	10,642	45,919	27,229	24,905	23,109	15,899	4,763	5,382	0	1
98	13,294	3,690	5,694	23,458	11,285	33,795	9,260	8,155	8,723	28,814	25,977	8,952	37,345	22,705	20,650	19,059	13,176	3,792	4,561	36	-1
99	8,676	2,325	3,554	14,720	7,111	20,655	5,968	5,273	5,552	17,963	15,693	5,770	23,637	13,682	13,487	12,030	8,187	2,367	2,905	100	-1
100	15,168	4,264	6,597	27,532	12,907	39,464	10,769	9,262	9,938	33,724	29,431	10,204	43,686	26,168	23,565	21,882	15,127	4,513	5,128	8	0
101	15,734	4,409	6,808	28,480	13,318	40,563	11,217	9,546	10,262	34,980	30,343	10,496	44,992	26,766	24,450	22,743	15,641	4,676	5,294	4	0
102	13,137	3,633	5,589	23,191	11,061	33,408	9,109	7,966	8,516	28,332	25,384	8,797	36,479	22,510	20,328	18,705	12,901	3,761	4,459	38	-1
103	5,474	1,504	2,191	8,930	4,575	12,937	3,611	3,292	3,501	11,544	10,041	3,507	14,782	8,504	8,471	7,512	5,060	1,492	1,859	134	-1
104	13,629	3,790	5,953	24,130	11,744	35,176	9,599	8,398	9,035	29,920	26,713	9,311	38,658	23,334	21,360	19,823	13,681	3,945	4,722	28	-1
105	16,164	3,496	5,943	29,698	11,386	37,450	9,620	8,853	9,252	31,308	28,631	9,762	37,426	26,637	23,427	21,341	14,765	4,373	4,682	18	-1
106	17,934	3,856	6,680	32,994	12,761	42,733	10,475	9,686	10,193	34,764	32,387	10,881	42,066	29,606	25,992	23,626	16,310	4,873	5,199	0	1
107	10,695	2,278	3,807	18,349	7,684	24,841	5,997	5,602	5,826	20,437	19,018	6,160	24,588	17,226	14,948	13,510	9,324	2,835	3,044	78	-1
108	11,051	2,355	4,000	19,176	8,019	25,943	6,199	5,870	6,135	21,475	19,931	6,407	25,566	18,089	15,629	14,199	9,844	2,940	3,192	74	-1
109	14,146	3,009	5,166	26,039	9,888	32,259	8,342	7,799	8,120	27,410	24,417	8,503	32,548	23,616	20,240	18,435	12,842	3,832	4,086	40	-1
110	17,051	3,631	6,200	31,045	11,965	39,620	9,948	9,233	9,691	32,935	30,279	10,198	39,097	27,955	24,669	22,407	15,466	4,590	4,902	10	0
111	17,747	3,802	6,512	32,621	12,466	41,695	10,379	9,569	10,066	34,239	31,656	10,697	41,035	29,143	25,719	23,298	16,089	4,813	5,101	2	0
112	6,580	1,538	2,436	11,822	4,841	15,612	3,807	3,629	3,860	13,214	12,293	3,959	15,738	11,294	9,399	8,632	6,043	1,776	1,993	120	-1
113	8,606	2,376	3,226	15,564	6,441	22,025	5,361	4,607	4,980	17,040	16,403	5,431	21,911	13,999	13,258	11,671	7,760	2,521	2,523	84	-1
114	16,649	4,402	6,446	30,551	12,562	44,521	9,789	8,872	9,768	34,198	32,834	10,476	42,343	27,316	25,895	23,293	15,422	5,122	4,800	4	0
115	16,251	4,336	6,339	29,810	12,374	43,695	9,547	8,694	9,576	33,552	32,432	10,219	41,569	26,896	25,311	22,795	15,109	5,010	4,710	6	0
116	4,121	1,138	1,766	7,046	3,586	10,851	2,718	2,439	2,678	9,333	8,678	2,631	11,839	7,155	6,749	6,036	4,143	1,247	1,408	130	-1
117	1,416	0,376	0,631	2,919	1,137	4,412	0,919	0,726	0,838	3,097	2,940	0,967	4,256	2,535	2,303	2,037	1,406	0,483	0,414	160	-1

Appendix A2. Learning and testing data

N	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19	X20	Stat
118	2,989	0,765	1,230	5,571	2,362	8,091	1,755	1,665	1,881	6,591	6,667	1,844	7,901	5,161	4,867	4,491	3,079	0,881	0,985	142	-1
119	0,519	0,158	0,245	1,268	0,400	2,088	0,174	0,214	0,272	1,082	1,365	0,379	1,214	0,984	0,976	0,801	0,494	0,201	0,133	174	-1
120	17,355	4,571	6,633	31,248	13,132	45,930	10,088	9,131	10,050	35,464	34,384	10,701	43,777	27,818	26,778	24,059	15,912	5,206	5,008	0	1
121	15,814	3,667	5,268	27,033	11,121	38,055	8,456	7,632	8,142	29,292	28,217	9,068	36,075	21,113	22,800	20,095	13,390	4,276	4,307	30	-1
122	13,476	3,237	4,636	23,919	9,532	32,772	7,344	6,618	7,069	25,974	24,431	7,707	32,300	18,299	19,502	17,232	11,597	3,736	3,746	48	-1
123	18,898	4,398	6,335	32,924	13,161	45,432	10,149	9,228	9,785	34,807	33,222	10,987	42,771	26,279	27,026	23,766	15,943	5,107	5,193	0	1
124	18,599	4,334	6,233	32,517	12,948	44,920	10,013	9,034	9,599	34,144	32,790	10,872	42,169	25,708	26,644	23,435	15,742	5,028	5,109	2	0
125	11,320	2,707	3,860	19,710	8,127	27,840	6,116	5,532	5,865	21,580	20,529	6,522	27,936	14,817	16,397	14,282	9,699	3,144	3,146	68	-1
126	4,453	0,888	1,298	7,782	2,799	9,866	2,067	1,881	2,049	8,220	6,983	2,138	11,087	4,862	5,604	4,698	3,181	1,206	1,098	140	-1
127	4,035	0,803	1,221	7,086	2,561	8,953	1,935	1,747	1,888	7,500	6,262	1,984	10,030	4,561	5,052	4,359	2,942	1,084	1,022	142	-1
128	17,927	4,215	5,942	30,961	12,543	42,918	9,706	8,757	9,273	32,991	31,515	10,420	40,420	24,870	25,681	22,598	15,217	4,819	4,927	10	0
129	1,133	0,259	0,258	1,807	0,614	2,478	0,545	0,464	0,426	1,871	1,930	0,485	2,867	1,382	1,227	1,091	0,672	0,292	0,240	184	-1
130	6,151	1,513	2,226	10,334	4,647	15,044	3,884	3,289	3,544	11,845	11,216	3,832	14,624	9,912	8,996	8,275	5,576	1,702	1,840	124	-1
131	15,843	4,322	6,079	28,196	12,377	40,488	10,348	8,752	9,329	32,333	30,052	9,907	39,694	26,489	24,089	21,940	14,905	4,686	4,720	10	0
132	16,397	4,489	6,433	29,634	12,944	42,335	10,826	9,243	9,909	34,077	31,421	10,411	41,467	27,736	25,320	23,268	15,868	4,924	4,949	2	0
133	16,947	4,548	6,591	30,157	13,320	43,561	11,120	9,360	10,049	34,662	32,301	10,677	42,319	28,075	26,092	23,735	16,114	5,017	5,066	0	1
134	3,389	0,852	1,087	5,544	2,408	7,939	2,175	1,719	1,761	6,287	5,752	1,952	8,854	4,965	4,475	4,144	2,794	0,962	0,914	160	-1
135	10,272	2,624	3,764	16,990	8,032	26,154	6,313	5,502	5,877	20,127	19,735	6,338	24,294	16,215	15,607	14,199	9,544	2,854	3,022	74	-1
136	8,188	2,137	3,078	13,793	6,448	20,735	5,144	4,553	4,841	16,137	15,701	5,193	19,894	13,380	12,445	11,277	7,604	2,271	2,480	92	-1
137	8,753	2,362	3,503	17,384	6,575	22,925	5,445	4,883	5,400	18,010	18,215	5,552	24,474	15,002	13,170	12,344	8,220	2,763	2,602	90	-1
138	12,733	3,461	5,230	25,345	9,791	33,859	7,863	7,227	7,954	27,462	25,843	8,015	35,708	22,103	19,521	18,457	12,338	4,141	3,826	48	-1
139	12,951	3,522	5,312	25,756	9,949	34,350	8,074	7,326	8,079	28,022	26,270	8,103	36,401	22,666	19,764	18,704	12,523	4,235	3,875	46	-1
140	10,276	2,741	4,044	20,189	7,651	26,625	6,265	5,657	6,222	21,283	20,994	6,326	28,229	17,641	15,377	14,341	9,540	3,260	3,002	74	-1
141	16,582	4,463	6,796	32,607	12,762	43,758	10,590	9,346	10,211	36,039	32,505	10,326	47,278	28,375	25,114	23,607	15,879	5,450	4,899	4	0
142	16,849	4,547	6,958	33,177	13,044	44,810	10,802	9,540	10,417	36,633	33,542	10,584	48,089	28,796	25,722	24,172	16,236	5,555	5,013	0	1
143	16,233	4,379	6,691	31,834	12,591	42,894	10,480	9,184	10,034	35,227	31,949	10,191	46,414	27,837	24,562	23,210	15,607	5,333	4,830	8	0
144	13,727	3,767	5,739	27,227	10,724	36,747	8,703	7,874	8,644	29,946	28,136	8,654	39,661	24,234	20,818	19,875	13,309	4,526	4,136	32	-1
145	15,675	4,087	6,004	28,884	11,896	39,503	9,287	8,325	9,027	32,757	30,162	9,109	40,386	24,925	23,129	21,066	14,230	4,692	4,656	20	-1
146	9,712	2,401	3,712	17,543	7,423	25,498	5,407	5,000	5,366	20,517	18,133	5,551	25,613	15,291	13,934	12,717	8,640	3,002	2,756	78	-1
147	12,058	3,135	4,774	22,509	9,328	31,004	7,100	6,515	7,015	25,725	23,441	7,063	31,544	18,900	18,097	16,511	11,140	3,623	3,634	54	-1

Appendix A2. Learning and testing data

N	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19	X20	Stat
148	13,633	3,529	5,336	25,319	10,463	34,784	8,153	7,365	7,944	28,884	27,145	7,929	35,041	22,161	20,284	18,660	12,658	4,024	4,141	38	-1
149	7,376	1,817	2,821	13,503	5,639	19,738	3,987	3,743	4,121	15,783	14,385	4,220	18,325	11,929	11,069	10,022	6,745	2,284	2,132	100	-1
150	17,570	4,476	6,664	32,032	13,271	44,554	10,186	9,156	10,011	36,488	33,585	10,133	45,364	27,540	25,686	23,484	15,840	5,245	5,112	2	0
151	17,315	4,431	6,617	31,678	13,152	43,978	10,056	9,103	9,959	36,131	33,281	10,039	45,013	27,126	25,454	23,314	15,724	5,181	5,081	4	0
152	17,626	4,495	6,684	32,149	13,312	44,702	10,204	9,189	10,047	36,616	33,716	10,161	45,453	27,639	25,801	23,563	15,892	5,262	5,128	0	1
153	13,548	3,935	6,063	27,546	11,018	39,185	9,100	7,800	8,509	29,682	28,589	9,359	37,797	24,466	21,824	20,599	13,931	4,486	4,323	22	-1
154	10,908	3,248	4,867	22,145	8,950	32,029	7,353	6,175	6,802	24,167	23,282	7,436	30,914	19,452	17,561	16,505	11,155	3,693	3,426	46	-1
155	2,161	0,691	1,010	4,502	1,801	6,432	1,548	1,282	1,424	4,626	4,960	1,577	6,429	3,764	3,707	3,209	2,058	0,722	0,731	144	-1
156	4,995	1,462	2,145	9,917	3,980	14,337	3,251	2,892	3,180	10,681	11,020	3,468	12,654	9,006	8,370	7,691	5,073	1,617	1,618	114	-1
157	15,797	4,522	7,060	32,775	12,626	45,515	10,661	9,100	9,864	34,655	33,442	10,880	44,157	28,418	25,593	23,864	16,214	5,276	4,992	0	1
158	15,595	4,462	6,938	32,328	12,432	44,942	10,441	8,954	9,684	34,061	32,911	10,723	43,510	27,853	25,305	23,484	15,951	5,193	4,919	2	0
159	14,545	4,221	6,453	30,000	11,630	41,855	9,874	8,370	9,095	31,894	30,687	9,995	40,504	26,153	23,477	22,093	14,964	4,852	4,611	12	-1
160	15,037	4,295	6,658	31,039	11,918	42,823	10,089	8,614	9,336	32,852	31,609	10,215	41,835	26,888	24,080	22,608	15,275	4,970	4,745	8	0
161	16,453	5,124	5,975	27,715	12,502	35,608	11,482	9,713	9,924	30,303	32,118	10,354	42,259	26,305	24,421	21,366	14,025	4,177	5,378	6	0
162	17,521	5,352	6,275	29,015	13,252	37,996	12,073	10,203	10,372	31,761	34,246	10,959	44,516	27,707	25,868	22,585	14,789	4,426	5,653	0	1
163	16,033	4,949	5,746	26,559	12,186	34,205	11,138	9,471	9,645	29,236	30,962	10,059	40,778	25,490	23,741	20,652	13,591	3,992	5,238	10	0
164	0,481	0,246	0,167	0,560	0,444	0,684	0,472	0,387	0,359	0,767	1,419	0,316	0,936	0,816	0,967	0,724	0,433	0,098	0,225	146	-1
165	3,880	1,196	1,372	6,135	2,982	8,624	2,835	2,132	2,140	7,239	7,845	2,239	11,246	5,705	5,405	4,657	3,010	1,013	1,195	120	-1
166	10,213	3,370	3,693	17,969	7,631	21,786	7,121	6,101	6,284	19,541	20,592	6,147	27,234	17,186	14,619	13,079	8,584	2,626	3,371	58	-1
167	14,123	4,435	5,087	23,662	10,761	30,180	9,700	8,392	8,572	26,311	27,326	8,773	36,514	22,509	20,738	18,357	12,059	3,537	4,637	24	-1
168	11,250	3,643	4,147	19,748	8,494	24,168	7,778	6,748	6,961	21,585	22,281	6,849	30,241	18,745	16,110	14,535	9,503	2,888	3,719	50	-1
169	16,747	4,353	6,837	32,729	12,553	41,138	11,530	9,792	10,502	34,174	31,169	11,121	43,270	28,411	25,438	23,960	16,235	4,992	5,223	0	1
170	15,908	4,105	6,363	30,989	11,763	39,117	10,885	9,193	9,890	31,858	29,382	10,636	39,714	26,937	24,130	22,736	15,385	4,736	4,899	10	0
171	16,223	4,253	6,547	31,626	12,100	40,003	11,091	9,440	10,119	32,762	30,141	10,829	41,291	27,627	24,554	23,197	15,693	4,865	5,016	6	0
172	1,931	0,518	0,864	4,789	1,305	4,546	1,357	1,325	1,441	4,408	4,162	1,369	4,872	3,281	3,451	3,232	2,247	0,581	0,669	162	-1
173	7,338	1,953	3,055	14,268	5,679	19,012	5,182	4,321	4,693	14,907	15,030	5,100	18,514	13,024	11,558	10,929	7,343	2,136	2,407	104	-1
174	5,146	1,388	2,125	10,222	3,910	12,927	3,645	3,009	3,285	10,451	10,280	3,492	12,645	9,213	8,108	7,632	5,156	1,479	1,675	130	-1
175	8,020	2,132	3,264	15,297	6,188	20,562	5,618	4,652	5,054	16,098	15,821	5,531	20,296	14,178	12,438	11,683	7,849	2,339	2,581	98	-1
176	8,550	2,254	3,426	16,101	6,536	21,476	6,015	4,941	5,319	16,764	16,650	5,867	21,596	15,149	12,962	12,154	8,198	2,458	2,724	94	-1
177	5,900	1,710	2,755	12,770	4,719	15,640	4,214	3,692	4,034	12,651	11,311	4,241	16,547	11,601	8,872	8,576	6,116	1,701	2,034	118	-1

Appendix A2. Learning and testing data

Nr	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19	X20	Stat
178	11,302	3,060	4,640	23,262	8,269	28,498	7,259	6,538	7,007	22,725	20,552	7,532	30,162	19,452	16,358	15,371	10,722	3,203	3,541	58	-1
179	0,857	0,260	0,548	2,371	0,788	2,655	0,705	0,676	0,731	2,485	2,007	0,671	2,605	1,910	1,744	1,610	1,194	0,329	0,362	186	-1
180	16,899	4,235	6,592	32,541	12,266	41,580	10,822	9,321	9,924	32,959	30,934	10,755	44,251	28,096	23,663	22,011	15,271	4,694	5,069	8	0
181	17,123	4,304	6,719	33,455	12,361	42,227	11,001	9,471	10,074	33,510	31,672	10,886	44,895	28,701	23,910	22,393	15,514	4,819	5,111	6	0
182	16,643	4,177	6,483	32,031	12,066	40,748	10,710	9,181	9,771	32,457	30,281	10,568	43,593	27,695	23,217	21,620	15,014	4,615	4,995	12	-1
183	0,158	0,043	0,097	0,521	0,098	0,400	0,126	0,127	0,129	0,366	0,303	0,126	0,376	0,475	0,244	0,264	0,196	0,050	0,064	198	-1
184	17,577	4,455	7,059	34,752	12,891	44,238	11,395	9,821	10,539	35,010	33,404	11,341	46,017	29,913	25,133	23,625	16,347	5,019	5,334	0	1
185	0,813	0,267	0,398	1,942	0,637	1,984	0,786	0,556	0,578	1,698	2,052	0,604	2,294	1,540	1,478	1,324	0,891	0,225	0,300	172	-1
186	16,267	5,064	7,059	34,621	12,793	46,219	11,353	9,165	9,930	34,406	35,944	10,919	46,436	28,109	25,721	23,567	15,857	5,217	5,110	0	1
187	15,718	4,802	6,748	33,010	12,311	44,608	10,808	8,728	9,437	32,922	34,016	10,505	44,543	26,869	24,542	22,431	15,085	5,047	4,844	8	0
188	15,320	4,678	6,578	31,972	12,038	43,081	10,573	8,596	9,302	32,218	32,983	10,249	43,524	26,211	23,929	21,946	14,753	4,890	4,757	12	-1
189	1,977	0,753	1,012	4,346	1,849	5,700	1,584	1,384	1,433	4,632	5,124	1,475	6,535	3,641	3,600	3,251	2,187	0,590	0,773	158	-1
190	14,979	4,503	6,387	30,759	11,794	42,108	10,209	8,349	9,014	31,077	31,873	10,053	42,221	25,649	23,181	21,278	14,329	4,755	4,623	16	-1
191	12,690	3,858	5,472	26,129	10,098	35,885	8,618	7,111	7,661	26,400	27,323	8,568	36,201	21,640	19,849	18,128	12,168	4,056	3,933	48	-1
192	15,799	4,828	6,780	33,187	12,369	44,775	10,876	8,782	9,499	33,126	34,155	10,558	44,730	27,010	24,709	22,570	15,183	5,069	4,873	6	0
193	17,680	4,275	6,235	30,748	12,645	42,787	11,179	8,991	9,427	32,532	32,330	10,673	45,144	27,035	24,327	22,348	15,096	4,769	4,986	8	0
194	18,864	4,533	6,512	32,313	13,431	45,184	11,640	9,469	9,916	34,541	34,098	11,171	47,635	28,707	25,631	23,482	15,868	5,027	5,255	0	1
195	17,612	4,255	6,220	30,606	12,615	42,678	11,152	8,960	9,391	32,393	32,254	10,643	45,037	26,933	24,209	22,284	15,048	4,753	4,970	10	0
196	15,227	3,625	5,296	26,393	10,752	36,839	9,444	7,623	7,967	27,563	27,893	9,111	38,017	22,621	21,028	19,206	12,811	4,060	4,224	36	-1
197	15,251	3,638	5,310	26,447	10,779	36,918	9,461	7,644	7,989	27,627	27,979	9,131	38,109	22,683	21,073	19,249	12,842	4,068	4,236	34	-1
198	6,048	1,468	2,269	10,848	4,409	15,836	3,946	3,056	3,229	11,032	10,875	3,954	17,252	9,262	7,919	7,775	5,193	1,767	1,664	128	-1
199	15,013	3,550	5,204	26,115	10,498	36,335	9,300	7,474	7,800	26,928	27,595	8,993	37,319	22,155	20,782	18,868	12,552	4,013	4,134	38	-1
200	12,735	3,123	4,714	23,398	9,009	31,236	8,302	6,666	6,960	23,522	24,251	7,939	33,343	19,124	17,749	16,700	11,072	3,421	3,681	58	-1

Testing data

N	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19	X20	Stat
201	16,753	4,286	5,966	29,648	11,868	39,062	9,876	8,506	9,145	31,018	28,953	10,031	43,607	26,134	22,640	20,120	13,413	4,356	4,895	18	-1
202	2,021	0,533	0,786	3,378	1,653	5,315	1,218	1,069	1,113	4,239	3,408	1,220	4,796	3,317	3,029	2,719	1,875	0,669	0,567	170	-1
203	8,011	2,104	3,070	13,634	6,136	19,679	4,971	4,208	4,547	15,338	15,046	4,922	21,069	13,041	11,088	10,180	6,652	2,173	2,385	96	-1
204	5,221	1,288	1,973	7,652	4,272	12,993	3,259	2,652	2,863	9,751	9,138	3,252	13,510	8,584	6,927	6,548	4,319	1,420	1,535	136	-1
205	18,207	4,701	6,560	32,770	12,897	42,379	10,768	9,401	10,195	34,162	31,891	10,950	46,750	28,883	24,999	22,330	14,891	4,842	5,355	4	0
206	17,689	4,570	6,397	31,515	12,612	41,258	10,481	9,140	9,878	33,187	31,354	10,598	45,804	28,252	24,103	21,651	14,416	4,663	5,222	10	0
207	18,603	4,879	6,800	33,554	13,391	44,149	11,063	9,613	10,464	35,344	32,927	11,264	48,269	30,035	25,680	23,024	15,369	5,015	5,496	0	1
208	13,266	3,386	4,628	22,448	9,539	30,207	7,769	6,698	7,227	24,973	22,694	7,600	34,034	20,529	17,986	15,745	10,527	3,493	3,812	48	-1
209	6,752	1,672	2,791	13,041	5,294	17,051	4,021	3,912	4,351	15,201	12,506	4,200	16,469	11,948	11,062	9,920	7,006	1,975	2,157	118	-1
210	17,187	4,215	6,761	29,927	13,575	41,356	10,474	9,777	10,501	36,180	28,894	10,611	43,765	28,051	26,202	23,253	16,165	4,832	5,348	0	1
211	17,060	4,200	6,733	29,734	13,510	41,170	10,406	9,714	10,435	35,968	28,781	10,544	43,608	27,919	26,006	23,111	16,063	4,807	5,319	2	0
212	1,229	0,270	0,697	2,750	1,190	3,610	0,897	0,873	1,014	3,410	2,605	0,911	3,507	2,682	2,405	2,237	1,679	0,389	0,506	182	-1
213	15,481	3,819	5,994	26,893	12,155	37,160	9,273	8,767	9,407	32,589	25,744	9,522	38,711	24,808	23,849	21,054	14,634	4,349	4,807	24	-1
214	13,042	2,955	4,933	22,632	9,874	30,484	7,839	7,242	7,712	26,745	21,150	7,810	32,539	20,661	19,346	17,043	11,885	3,674	3,889	52	-1
215	16,177	3,967	6,218	27,791	12,691	38,527	9,616	9,110	9,770	33,620	26,833	9,899	40,469	25,854	24,628	21,702	15,050	4,464	5,012	16	-1
216	16,504	4,060	6,445	28,672	13,031	39,795	9,847	9,384	10,078	34,708	27,734	10,168	41,803	26,867	25,200	22,381	15,580	4,633	5,145	10	0
217	13,036	3,373	5,104	23,787	10,027	35,304	7,616	6,837	7,243	25,355	23,657	8,512	35,058	21,543	18,081	17,532	11,805	3,713	3,783	46	-1
218	5,354	1,182	1,814	8,651	3,888	12,347	2,928	2,657	2,753	9,457	8,840	3,108	13,265	8,127	6,465	6,341	4,313	1,252	1,478	128	-1
219	17,764	4,509	6,767	30,632	13,673	46,058	10,355	9,244	9,750	34,668	31,612	10,999	48,167	27,746	24,416	22,936	15,285	5,024	5,079	2	0
220	17,271	4,385	6,576	29,985	13,232	44,901	10,124	8,947	9,441	33,525	30,567	10,766	46,800	27,049	23,561	22,247	14,809	4,910	4,914	6	0
221	17,995	4,540	6,927	31,598	13,786	46,634	10,482	9,420	9,932	35,320	32,103	11,137	48,757	28,308	24,764	23,366	15,578	5,138	5,151	0	1
222	13,635	3,480	5,320	24,602	10,488	36,558	7,920	7,125	7,540	26,662	24,639	8,710	37,286	22,442	18,607	18,007	12,113	3,956	3,908	42	-1
223	3,276	0,789	1,224	5,877	2,451	8,134	1,958	1,740	1,831	6,243	5,836	2,062	8,611	5,210	4,356	4,355	2,968	0,816	0,979	146	-1
224	2,395	0,535	0,938	4,584	1,703	5,905	1,514	1,282	1,339	4,688	4,307	1,474	6,876	3,807	3,112	3,104	2,107	0,655	0,687	156	-1
225	18,601	4,474	6,232	30,522	13,379	42,318	11,505	9,965	9,942	34,267	33,664	10,724	43,652	28,895	26,573	23,560	16,118	5,054	5,159	0	1
226	8,685	2,033	3,159	15,238	6,270	19,889	5,913	4,942	5,015	16,640	16,443	5,246	20,813	14,005	13,082	11,718	7,920	2,440	2,563	84	-1
227	0,849	0,222	0,317	1,505	0,614	2,189	0,536	0,457	0,427	1,608	1,601	0,495	1,900	1,533	1,317	1,054	0,703	0,264	0,225	156	-1

Appendix A2. Learning and testing data

Nr	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19	X20	Stat
228	18,164	4,348	6,122	30,072	13,003	41,245	11,181	9,752	9,791	33,530	33,107	10,460	43,043	28,359	25,785	22,980	15,664	4,921	5,074	2	0
229	0,613	0,163	0,250	1,166	0,457	1,694	0,402	0,353	0,323	1,301	1,224	0,357	1,433	1,211	1,004	0,795	0,536	0,221	0,165	162	-1
230	9,866	2,458	3,613	17,451	7,262	22,991	6,518	5,608	5,782	19,529	19,095	5,898	23,819	16,044	15,084	13,499	9,105	2,786	2,964	76	-1
231	15,690	3,832	5,355	25,471	11,461	35,744	9,895	8,428	8,512	29,085	29,280	9,072	38,107	24,057	22,287	19,930	13,421	4,236	4,422	26	-1
232	17,418	4,239	5,990	29,343	12,569	40,029	10,888	9,483	9,553	32,558	32,471	10,174	41,643	27,354	25,156	22,482	15,318	4,772	4,927	8	0
233	17,929	4,282	6,962	31,514	13,419	43,350	11,238	9,920	10,025	37,904	32,736	10,164	46,569	27,703	27,037	23,961	16,031	5,458	5,003	0	1
234	6,945	1,819	2,877	12,368	5,460	16,543	4,571	4,107	4,165	16,175	11,988	3,884	21,083	10,910	10,437	9,052	6,215	2,203	2,061	102	-1
235	9,396	2,230	3,571	15,881	7,018	21,571	5,895	5,191	5,311	20,587	16,298	4,995	25,725	14,048	13,911	11,998	8,120	2,847	2,633	82	-1
236	8,312	2,064	3,257	14,307	6,311	18,912	5,446	4,791	4,854	18,405	14,435	4,570	23,514	12,503	12,428	10,790	7,323	2,504	2,410	88	-1
237	10,982	2,676	4,199	19,021	8,217	25,382	7,012	6,128	6,299	23,574	19,192	6,107	29,107	17,314	16,381	14,219	9,696	3,238	3,155	68	-1
238	17,543	4,242	6,813	30,835	13,128	42,696	11,026	9,690	9,804	37,020	32,279	9,994	45,825	27,164	26,513	23,436	15,656	5,356	4,905	2	0
239	16,653	4,110	6,616	29,830	12,601	41,272	10,595	9,315	9,487	35,643	31,221	9,655	44,047	26,383	25,492	22,557	15,136	5,132	4,754	8	0
240	1,323	0,208	0,380	2,125	0,747	2,898	0,569	0,575	0,624	2,670	1,816	0,580	3,382	1,704	1,545	1,451	0,966	0,404	0,235	156	-1
241	19,559	4,565	6,067	32,827	13,014	42,847	11,424	9,505	10,301	36,744	32,475	10,381	44,326	28,482	26,638	23,854	15,892	5,122	5,242	0	1
242	15,921	3,515	4,786	26,388	10,301	34,948	9,152	7,520	8,106	29,344	26,379	8,266	36,006	22,577	21,030	18,933	12,499	4,183	4,128	28	-1
243	15,321	3,419	4,654	25,762	9,911	34,029	8,730	7,244	7,828	28,563	25,588	7,941	35,037	21,881	20,259	18,289	12,069	4,084	3,962	36	-1
244	1,226	0,225	0,244	1,574	0,706	2,483	0,562	0,429	0,414	1,822	1,716	0,502	2,187	1,392	1,397	1,074	0,714	0,258	0,255	188	-1
245	18,154	4,183	5,549	30,223	12,030	40,102	10,500	8,717	9,422	34,055	29,793	9,592	41,101	26,225	24,319	21,923	14,691	4,794	4,793	10	0
246	18,568	4,307	5,742	30,941	12,430	41,398	10,773	8,949	9,707	35,144	30,514	9,867	42,344	27,009	24,954	22,646	15,149	4,936	4,923	6	0
247	5,781	1,182	1,600	9,730	3,394	11,552	3,031	2,711	2,896	10,580	8,825	2,720	13,031	7,793	7,203	6,478	4,251	1,475	1,422	142	-1
248	9,746	2,026	2,725	16,150	5,857	20,009	5,132	4,479	4,779	17,895	15,367	4,572	21,396	13,700	12,165	10,951	7,182	2,493	2,406	96	-1
249	17,607	4,377	6,444	30,527	12,950	44,610	11,305	9,023	9,428	32,457	33,417	10,958	43,318	27,024	27,079	22,247	14,734	4,755	5,071	8	0
250	17,998	4,476	6,612	31,222	13,293	45,828	11,541	9,196	9,636	33,479	34,177	11,142	44,659	27,844	27,524	22,717	15,109	4,903	5,157	2	0
251	18,136	4,515	6,665	31,540	13,387	46,226	11,601	9,249	9,715	33,768	34,387	11,228	45,026	28,049	27,756	22,875	15,216	4,952	5,193	0	1
252	16,644	4,100	6,086	28,780	12,257	42,443	10,605	8,551	8,898	30,572	31,588	10,433	41,234	25,227	25,684	21,081	14,011	4,544	4,756	16	-1
253	0,746	0,158	0,209	0,983	0,559	2,198	0,402	0,257	0,249	1,417	1,214	0,349	1,316	1,103	1,104	0,891	0,608	0,255	0,136	172	-1
254	16,602	4,093	6,078	28,735	12,232	42,350	10,583	8,542	8,889	30,507	31,531	10,420	41,166	25,180	25,633	21,047	13,990	4,534	4,750	18	-1
255	10,723	2,650	4,053	18,718	7,996	27,064	7,059	5,552	5,812	20,147	20,437	6,672	26,571	16,929	16,354	13,710	9,150	2,987	3,115	72	-1
256	4,903	1,184	1,658	8,350	3,434	11,178	3,020	2,515	2,597	8,551	9,197	2,889	11,463	7,252	7,088	6,064	4,023	1,214	1,406	138	-1
257	11,171	3,079	4,205	19,769	8,613	27,648	6,952	6,174	6,640	22,933	20,211	7,006	26,226	17,165	18,154	15,975	10,677	3,262	3,267	70	-1

Appendix A2. Learning and testing data

Nr	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19	X20	Stat
258	2,160	0,547	0,862	3,689	1,797	4,987	1,185	1,289	1,393	4,970	4,631	1,167	4,769	3,377	3,555	3,356	2,131	0,685	0,700	164	-1
259	6,808	1,694	2,155	10,742	4,834	14,523	3,954	3,557	3,715	12,429	11,290	3,888	14,019	9,225	10,170	8,871	5,747	1,815	1,886	120	-1
260	12,475	3,441	4,858	22,565	9,710	31,365	7,785	6,909	7,509	26,243	23,218	7,819	29,091	19,625	20,545	18,225	12,134	3,652	3,718	52	-1
261	17,676	4,719	6,637	31,583	13,320	42,805	10,519	9,568	10,308	36,090	31,663	10,744	41,550	26,677	28,089	24,532	16,169	4,989	5,162	0	1
262	8,310	2,148	2,827	13,397	6,191	19,134	4,960	4,367	4,672	15,885	13,963	4,954	17,959	12,091	12,852	11,103	7,294	2,304	2,374	106	-1
263	17,203	4,577	6,566	30,939	13,065	42,160	10,275	9,385	10,098	35,282	31,238	10,554	41,076	26,063	27,445	24,004	15,848	4,906	5,050	2	0
264	16,918	4,503	6,510	30,346	12,964	41,799	10,161	9,261	9,978	34,834	30,997	10,425	40,772	25,761	27,058	23,706	15,665	4,848	4,975	4	0
265	17,826	4,404	6,321	28,823	13,672	40,812	10,413	9,564	10,086	35,511	32,472	9,919	42,708	28,223	25,975	22,738	15,407	4,953	5,180	4	0
266	17,869	4,419	6,332	28,873	13,708	40,899	10,448	9,586	10,106	35,575	32,584	9,942	42,813	28,279	26,035	22,785	15,441	4,958	5,195	2	0
267	12,802	3,103	4,388	20,302	9,696	28,411	7,341	6,768	7,178	25,251	22,957	6,924	30,240	20,035	18,130	16,055	10,864	3,532	3,684	52	-1
268	2,929	0,673	0,948	4,364	2,228	6,428	1,614	1,441	1,463	5,953	5,210	1,327	6,786	4,476	4,018	3,369	2,318	0,874	0,762	156	-1
269	6,968	1,619	2,152	11,314	4,835	14,592	3,686	3,511	3,666	13,411	12,144	3,398	15,982	10,435	9,521	8,257	5,627	1,841	1,856	108	-1
270	6,068	1,386	1,923	9,777	4,326	12,590	3,276	3,148	3,278	11,950	11,002	2,900	14,307	9,049	8,374	7,359	5,036	1,619	1,653	114	-1
271	10,488	2,421	3,377	16,498	7,541	22,222	5,872	5,395	5,670	19,734	18,058	5,481	25,010	16,273	14,018	12,407	8,409	2,764	2,912	74	-1
272	18,146	4,494	6,473	29,372	13,967	41,873	10,628	9,768	10,281	36,182	33,051	10,209	44,074	28,803	26,344	23,137	15,699	5,079	5,299	0	1
273	8,406	2,466	3,352	15,959	6,528	21,099	5,094	4,729	5,187	18,789	15,459	5,052	21,699	12,840	13,170	12,266	8,110	2,658	2,491	112	-1
274	1,379	0,439	0,609	2,839	1,114	3,615	0,769	0,905	0,953	3,224	2,508	0,921	3,585	2,013	2,392	2,216	1,453	0,527	0,444	184	-1
275	8,453	2,476	3,376	16,048	6,564	21,241	5,128	4,753	5,214	18,819	15,567	5,102	21,796	12,920	13,252	12,341	8,155	2,666	2,508	110	-1
276	16,864	4,502	6,526	32,300	12,384	40,682	10,657	9,265	9,984	35,705	30,933	9,903	44,292	26,360	25,269	22,832	15,333	4,969	4,930	10	0
277	16,956	4,518	6,549	32,385	12,456	40,865	10,704	9,305	10,033	35,813	31,085	9,963	44,422	26,487	25,416	22,955	15,416	4,987	4,958	8	0
278	17,874	4,663	6,854	33,578	13,169	43,192	11,209	9,712	10,480	37,301	32,220	10,571	46,208	27,758	26,749	23,973	16,123	5,197	5,215	0	1
279	12,554	3,394	4,611	23,093	9,100	29,496	7,270	6,699	7,307	26,061	22,637	7,088	31,929	18,995	18,260	16,635	10,984	3,584	3,609	62	-1
280	2,104	0,641	0,844	3,695	1,703	5,156	1,254	1,285	1,350	4,484	3,614	1,361	5,766	2,883	3,209	3,042	2,007	0,711	0,632	176	-1
281	13,395	4,311	5,708	24,724	11,196	34,780	10,049	8,268	8,506	27,090	25,366	9,872	37,158	23,095	21,366	19,796	13,307	4,172	4,422	26	-1
282	15,225	4,698	6,348	27,809	12,462	38,762	11,094	9,211	9,462	30,184	28,384	10,971	41,787	25,644	23,721	21,976	14,719	4,641	4,943	10	0
283	16,270	5,026	6,854	29,922	13,398	41,311	11,872	9,996	10,322	32,865	30,642	11,706	44,074	28,154	25,710	23,807	16,033	4,973	5,354	0	1
284	15,739	4,899	6,679	29,045	13,039	40,308	11,604	9,680	9,991	31,682	29,714	11,467	43,320	26,869	24,952	23,061	15,542	4,835	5,206	4	0
285	9,452	3,113	3,992	17,321	7,958	24,400	6,997	5,991	6,133	19,126	18,335	7,018	25,211	16,091	15,710	14,358	9,628	2,880	3,210	66	-1
286	5,562	1,730	2,192	9,556	4,574	15,115	3,880	3,208	3,268	9,809	10,873	4,242	13,979	9,109	8,859	8,086	5,373	1,623	1,747	124	-1
287	8,333	2,791	3,568	15,482	7,094	22,120	6,127	5,270	5,431	16,884	16,567	6,280	21,955	14,375	14,066	12,904	8,654	2,539	2,870	82	-1

Appendix A2. Learning and testing data

N	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19	X20	Stat
288	8,806	2,919	3,721	16,057	7,459	22,888	6,492	5,598	5,735	17,665	17,495	6,595	23,115	15,094	14,772	13,541	9,070	2,647	3,031	76	-1
289	1,491	0,242	0,369	2,596	0,699	2,837	0,550	0,607	0,685	2,418	2,196	0,636	2,357	2,454	1,776	1,368	0,861	0,341	0,346	168	-1
290	17,506	5,016	6,798	31,717	13,438	42,291	11,479	10,100	10,540	34,864	35,191	10,717	46,020	29,783	25,856	23,188	15,459	4,798	5,471	0	1
291	16,573	4,695	6,471	29,378	12,859	39,378	10,984	9,642	10,047	32,913	33,380	10,102	44,261	28,289	24,129	21,725	14,478	4,488	5,234	6	0
292	9,462	2,723	3,799	16,610	7,555	22,116	6,478	5,739	5,994	19,072	19,973	5,792	25,631	17,443	13,695	12,518	8,413	2,479	3,122	78	-1
293	13,495	3,852	5,233	23,797	10,420	30,686	9,069	8,113	8,368	26,925	27,053	8,176	35,865	23,633	19,585	17,593	11,846	3,557	4,379	42	-1
294	4,417	1,204	1,801	8,679	3,231	10,589	2,879	2,611	2,777	8,686	9,027	2,883	11,968	7,775	6,369	5,761	3,858	1,132	1,479	134	-1
295	2,704	0,650	0,960	4,991	1,744	6,126	1,596	1,380	1,517	4,976	5,346	1,508	6,456	5,018	3,522	2,991	1,954	0,695	0,805	154	-1
296	16,892	4,750	6,602	30,433	12,964	40,076	11,122	9,804	10,238	33,740	33,765	10,248	44,798	28,830	24,620	22,228	14,803	4,596	5,303	4	0
297	15,076	4,313	6,364	27,827	12,555	40,494	9,704	8,876	9,368	31,102	29,551	10,421	38,857	26,221	24,381	22,117	15,105	4,446	4,866	10	0
298	8,570	2,517	3,770	16,247	7,283	24,167	5,626	5,042	5,426	18,038	17,116	6,033	22,330	15,992	14,101	12,753	8,783	2,633	2,760	88	-1
299	15,883	4,520	6,736	29,079	13,330	43,298	10,347	9,249	9,793	32,625	31,035	11,111	40,777	27,256	25,847	23,369	15,965	4,727	5,108	0	1
300	3,091	0,694	1,241	5,485	2,420	8,777	1,824	1,477	1,607	5,797	5,716	2,054	7,160	5,391	4,556	4,176	2,781	0,950	0,859	148	-1
301	9,228	2,671	4,051	17,514	7,822	26,376	5,917	5,368	5,821	19,538	18,596	6,522	23,811	17,102	15,243	13,820	9,504	2,857	2,957	70	-1
302	15,720	4,481	6,654	28,803	13,164	42,796	10,226	9,145	9,652	32,081	30,725	11,011	40,353	26,959	25,485	23,058	15,755	4,654	5,055	4	0
303	2,053	0,445	0,822	3,899	1,599	6,062	1,272	0,924	1,070	4,106	3,871	1,344	4,751	3,494	3,102	2,896	2,006	0,656	0,552	158	-1
304	12,414	3,484	5,067	22,175	10,279	33,434	7,621	6,952	7,449	25,442	24,060	8,263	31,201	20,928	19,958	17,792	12,101	3,622	3,888	42	-1
305	16,531	4,396	6,533	29,537	13,096	40,707	11,217	9,496	10,289	33,056	32,841	10,648	42,601	28,536	25,479	23,066	15,768	4,761	5,155	0	1
306	16,142	4,231	6,282	28,031	12,753	39,213	10,984	9,167	9,847	31,606	31,889	10,266	41,394	27,697	24,502	22,077	15,035	4,572	4,996	4	0
307	13,596	3,489	5,243	24,180	10,494	33,307	8,853	7,506	8,169	26,757	26,978	8,444	35,145	23,523	20,303	18,157	12,377	3,880	4,155	36	-1
308	13,733	3,530	5,300	24,440	10,614	33,660	8,937	7,610	8,282	27,131	27,201	8,542	35,508	23,726	20,594	18,434	12,566	3,933	4,195	32	-1
309	7,685	1,998	2,937	13,931	5,792	18,990	4,820	4,065	4,436	15,073	14,827	4,632	20,523	13,087	11,133	9,963	6,707	2,208	2,283	98	-1
310	3,699	0,872	1,520	6,365	3,000	9,237	2,368	2,097	2,254	7,320	7,611	2,307	10,006	6,760	5,332	5,173	3,418	1,107	1,153	138	-1
311	3,246	0,774	1,353	5,783	2,585	8,096	2,104	1,854	1,987	6,504	6,786	1,989	8,716	6,110	4,698	4,589	2,982	0,993	0,995	146	-1
312	15,902	4,154	6,197	27,727	12,541	38,814	10,767	8,991	9,686	31,199	31,410	10,098	40,956	27,312	24,086	21,686	14,779	4,532	4,905	8	0
313	17,249	4,292	6,416	30,911	12,897	40,670	11,572	9,841	10,317	31,978	31,877	11,511	43,214	25,291	25,780	23,538	16,176	4,424	5,420	0	1
314	16,943	4,182	6,311	30,341	12,654	40,078	11,350	9,568	10,106	31,413	31,343	11,293	42,134	24,865	25,306	23,189	15,878	4,339	5,329	2	0
315	0,533	0,150	0,248	1,254	0,400	1,534	0,408	0,350	0,380	1,016	1,239	0,442	1,636	0,744	0,920	0,814	0,556	0,125	0,192	166	-1
316	4,179	0,903	1,418	7,912	2,778	9,257	2,606	2,210	2,263	7,133	7,111	2,539	9,030	6,484	6,019	5,086	3,598	0,911	1,213	134	-1
317	3,539	0,780	1,186	6,795	2,307	7,667	2,067	1,901	1,938	6,271	6,270	2,006	7,970	5,565	4,876	4,232	2,969	0,761	1,017	140	-1

Appendix A2. Learning and testing data

Nr	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19	X20	Stat
318	9,566	2,391	3,373	17,743	6,694	22,433	6,135	5,058	5,302	17,158	16,950	6,041	22,594	14,406	13,632	12,367	8,491	2,439	2,821	88	-1
319	14,625	3,457	5,251	25,667	10,665	33,622	9,627	8,058	8,545	26,646	25,736	9,500	35,902	21,216	21,435	19,267	13,289	3,725	4,440	32	-1
320	16,265	4,008	5,992	29,267	11,987	37,978	10,827	9,131	9,694	30,220	29,623	10,704	40,563	23,864	24,120	22,034	15,123	4,179	5,066	8	0
321	6,237	1,937	2,942	13,176	5,296	17,982	4,233	4,051	4,339	13,613	13,914	4,749	16,943	11,535	10,296	10,218	6,873	2,063	2,070	108	-1
322	15,646	4,278	6,652	30,866	12,237	41,900	9,889	9,259	9,905	32,762	31,282	10,732	39,713	27,508	24,655	23,429	15,747	4,904	4,860	10	0
323	16,190	4,371	6,835	31,795	12,602	43,022	10,183	9,511	10,187	33,896	32,171	10,976	40,925	28,200	25,416	24,102	16,192	5,073	4,995	0	1
324	13,498	3,810	5,841	26,876	10,782	36,570	8,608	8,069	8,650	28,808	27,094	9,297	34,940	24,067	21,410	20,410	13,763	4,334	4,207	34	-1
325	14,190	3,936	6,095	28,085	11,248	38,720	8,934	8,359	8,965	29,960	28,661	9,767	36,395	24,924	22,437	21,386	14,411	4,510	4,413	28	-1
326	7,217	2,177	3,283	14,854	5,994	20,483	4,739	4,557	4,885	15,483	15,483	5,378	19,610	13,203	11,570	11,381	7,652	2,345	2,333	92	-1
327	5,977	1,843	2,852	12,902	5,054	17,389	4,113	3,879	4,168	13,181	13,249	4,595	16,442	11,310	9,792	9,823	6,649	2,013	1,977	110	-1
328	15,940	4,309	6,713	31,254	12,393	42,424	10,010	9,340	10,006	33,272	31,569	10,825	40,213	27,827	24,971	23,660	15,898	4,996	4,908	8	0
329	4,002	1,257	1,678	7,324	3,456	10,084	3,029	2,611	2,745	7,878	7,630	2,991	9,847	7,394	7,053	5,934	4,114	1,131	1,362	118	-1
330	6,252	1,643	2,294	10,387	4,933	14,765	4,184	3,598	3,756	11,372	11,096	4,110	14,626	10,960	9,755	8,000	5,524	1,576	1,935	106	-1
331	12,522	3,457	4,667	21,352	9,852	31,003	8,138	6,963	7,379	23,067	23,512	8,276	30,884	20,746	19,511	16,643	11,211	3,369	3,892	44	-1
332	9,117	2,516	3,475	15,459	7,291	22,974	6,051	5,157	5,432	16,677	16,993	6,247	22,137	15,446	14,687	12,255	8,282	2,433	2,850	72	-1
333	16,756	4,555	6,027	29,412	12,497	41,493	10,535	8,937	9,481	30,957	30,699	10,690	39,703	27,662	25,533	21,869	14,650	4,668	4,918	10	0
334	17,821	4,965	6,448	31,792	13,307	44,171	11,315	9,563	10,154	33,400	32,488	11,374	43,212	29,777	27,082	23,214	15,655	4,997	5,267	0	1
335	17,383	4,714	6,218	30,372	12,928	42,591	10,890	9,236	9,796	32,044	31,515	10,997	41,542	28,861	25,982	22,413	15,018	4,790	5,100	4	0
336	12,351	3,374	4,610	21,138	9,680	30,655	8,038	6,859	7,268	22,638	23,184	8,193	30,308	20,541	19,295	16,382	11,053	3,328	3,824	46	-1
337	4,334	1,128	1,582	7,383	3,333	11,641	2,599	2,241	2,388	8,238	8,251	2,837	10,175	6,855	6,434	6,003	4,052	1,230	1,265	126	-1
338	14,425	4,090	5,827	28,592	10,935	38,911	9,124	8,089	8,641	28,994	29,502	9,745	36,441	24,298	22,612	21,020	14,161	4,268	4,473	18	-1
339	12,209	3,513	5,134	24,663	9,480	34,062	8,027	6,923	7,449	24,929	24,670	8,636	30,542	20,908	19,687	18,442	12,467	3,677	3,874	36	-1
340	16,533	4,708	6,746	32,477	12,798	44,315	10,485	9,418	10,013	33,853	32,673	11,156	42,499	27,899	25,673	24,203	16,461	4,923	5,174	0	1
341	15,452	4,399	6,332	30,703	11,881	41,796	9,898	8,805	9,399	31,444	31,255	10,550	39,906	26,294	24,127	22,694	15,383	4,575	4,857	8	0
342	16,334	4,668	6,628	32,218	12,559	43,759	10,375	9,298	9,874	33,417	32,336	11,015	41,967	27,427	25,365	23,911	16,278	4,860	5,099	2	0
343	1,381	0,344	0,490	2,464	0,965	3,995	0,640	0,632	0,681	2,425	2,959	0,885	3,275	2,070	1,893	1,808	1,147	0,355	0,401	152	-1
344	10,361	2,974	4,411	20,960	8,088	28,593	6,991	5,979	6,444	21,546	20,526	7,360	25,792	18,145	16,786	15,744	10,678	3,138	3,315	58	-1
345	15,786	4,053	6,181	28,369	12,091	38,096	10,572	9,149	9,850	31,391	31,492	10,218	37,251	26,258	24,983	23,117	15,764	4,298	4,944	6	0
346	15,266	3,972	6,075	27,687	11,843	37,200	10,289	8,976	9,668	30,867	30,740	9,952	36,719	25,473	24,342	22,651	15,480	4,212	4,810	10	0
347	15,070	3,905	6,014	27,387	11,691	36,880	10,183	8,861	9,540	30,400	30,366	9,877	36,169	25,073	24,144	22,404	15,318	4,178	4,743	12	-1

Appendix A2. Learning and testing data

Nr	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19	X20	Stat
348	12,753	3,273	5,104	23,440	9,874	31,477	8,590	7,579	8,115	25,856	25,636	8,456	30,305	21,155	20,753	19,072	13,123	3,607	4,000	38	-1
349	14,492	3,734	5,813	26,310	11,294	35,435	9,752	8,600	9,279	29,293	29,456	9,549	34,494	24,286	23,311	21,761	14,871	3,995	4,609	18	-1
350	12,861	3,293	5,138	23,622	9,940	31,735	8,620	7,623	8,168	26,097	25,820	8,497	30,545	21,296	20,904	19,210	13,211	3,640	4,024	36	-1
351	2,362	0,685	1,119	5,637	1,797	6,317	1,706	1,614	1,687	5,132	5,390	1,794	6,742	3,629	4,277	4,067	2,812	0,715	0,817	154	-1
352	16,560	4,279	6,566	30,313	12,648	40,447	11,104	9,615	10,349	32,838	33,192	10,895	39,435	27,375	26,375	24,370	16,540	4,493	5,229	0	1
353	16,098	4,653	7,119	33,130	12,834	43,234	11,381	9,515	10,430	35,285	32,390	11,026	44,477	29,516	24,777	23,639	16,056	4,721	5,349	0	1
354	14,630	4,350	6,525	30,836	11,646	39,934	10,339	8,618	9,479	32,238	29,428	10,128	41,533	26,905	22,352	21,401	14,576	4,399	4,860	14	-1
355	2,639	0,918	1,279	5,828	2,270	8,519	2,140	1,380	1,546	5,860	5,073	2,055	7,841	4,756	4,066	3,925	2,648	0,861	0,839	142	-1
356	7,830	2,394	3,606	18,294	6,034	22,722	5,631	4,600	5,211	17,976	15,959	5,622	23,098	15,153	12,021	11,681	8,067	2,541	2,550	86	-1
357	11,641	3,390	5,099	24,574	9,026	31,812	8,128	6,727	7,410	24,903	22,741	8,229	32,399	21,066	17,685	16,752	11,440	3,476	3,817	48	-1
358	15,400	4,443	6,722	31,701	12,112	40,953	10,894	9,002	9,871	33,156	30,735	10,520	42,539	28,031	23,416	22,237	15,096	4,476	5,079	8	0
359	15,791	4,597	7,008	32,780	12,563	42,363	11,283	9,375	10,273	34,729	31,778	10,831	43,925	29,111	24,353	23,200	15,779	4,666	5,259	2	0
360	11,053	3,225	4,860	23,396	8,596	30,186	7,794	6,457	7,119	23,585	21,232	7,951	30,703	20,203	16,855	15,998	10,970	3,317	3,638	56	-1
361	17,802	4,200	6,774	30,660	13,597	40,324	11,393	10,198	10,547	35,434	31,884	10,638	47,023	28,170	25,224	23,102	15,690	4,880	5,337	0	1
362	6,722	1,685	2,362	10,832	5,019	14,459	4,259	3,682	3,680	12,890	10,860	3,776	18,093	10,399	8,684	7,901	5,327	1,830	1,863	124	-1
363	11,322	2,697	4,304	20,090	8,494	25,264	6,975	6,573	6,833	22,977	19,874	6,664	30,352	17,588	16,073	14,736	9,981	3,141	3,353	68	-1
364	17,183	4,068	6,613	29,647	13,245	39,046	11,090	9,988	10,284	34,402	31,115	10,373	45,567	26,930	24,572	22,692	15,393	4,714	5,194	4	0
365	8,392	2,050	3,206	14,542	6,404	18,578	5,316	4,917	4,995	17,046	14,302	4,929	22,926	13,422	11,671	10,730	7,298	2,306	2,474	98	-1
366	3,019	0,654	0,961	5,040	2,020	6,090	1,599	1,476	1,500	5,727	3,837	1,472	8,642	3,884	3,392	3,000	2,063	0,766	0,735	160	-1
367	7,660	1,933	2,863	13,409	5,721	16,948	4,839	4,416	4,465	15,391	13,296	4,415	21,200	12,286	10,403	9,522	6,450	2,125	2,211	108	-1
368	17,092	4,050	6,596	29,559	13,181	38,866	11,042	9,946	10,239	34,286	31,001	10,319	45,410	26,840	24,435	22,588	15,328	4,689	5,176	6	0
369	16,351	4,282	6,475	31,026	12,394	40,721	10,952	9,340	9,846	33,000	31,641	10,387	43,412	26,451	26,052	21,857	15,138	4,946	4,954	10	0
370	16,954	4,482	6,824	32,422	12,961	42,652	11,433	9,781	10,299	34,755	33,054	10,825	45,173	27,647	27,316	22,980	15,900	5,220	5,176	0	1
371	4,927	1,473	2,365	11,173	3,973	12,788	4,177	3,275	3,440	11,085	9,533	3,587	15,863	9,883	7,806	7,068	5,068	1,733	1,669	124	-1
372	6,544	1,828	2,797	13,868	4,920	16,215	5,140	3,930	4,146	13,815	12,733	4,288	19,086	11,898	10,107	8,921	6,218	2,182	2,017	110	-1
373	10,616	3,014	4,572	21,929	8,387	27,129	7,893	6,498	6,896	23,087	20,828	7,057	31,275	18,276	17,307	14,786	10,476	3,495	3,346	66	-1
374	1,567	0,474	0,735	3,846	1,170	3,996	1,289	0,998	1,076	3,659	2,855	1,069	4,214	2,917	2,901	2,373	1,632	0,573	0,472	168	-1
375	5,773	1,718	2,664	12,505	4,610	14,589	4,779	3,722	3,909	12,657	11,126	4,062	17,949	11,176	9,036	8,129	5,752	1,947	1,922	114	-1
376	16,520	4,313	6,521	31,195	12,525	40,963	11,053	9,446	9,939	33,247	31,900	10,482	43,797	26,653	26,276	22,041	15,272	4,973	5,008	6	0
377	12,481	3,011	4,539	20,784	9,237	27,780	8,418	6,699	6,747	22,303	19,864	7,661	30,943	19,137	16,688	15,093	10,281	3,376	3,468	64	-1

Appendix A2. Learning and testing data

N	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19	X20	Stat.
378	7,983	1,791	2,947	13,554	5,842	18,468	5,171	4,191	4,177	15,098	13,223	4,580	19,403	12,150	11,016	9,636	6,565	2,350	2,104	116	-1
379	16,520	4,186	6,417	30,932	12,083	38,951	11,131	9,324	9,514	31,170	28,623	10,577	40,596	27,141	23,724	21,794	14,727	4,776	4,785	14	-1
380	17,200	4,412	6,693	32,115	12,677	40,510	11,477	9,808	9,978	32,787	30,050	10,961	42,111	28,304	24,919	22,829	15,462	4,966	5,032	8	0
381	17,327	4,448	6,729	32,251	12,772	40,812	11,537	9,855	10,014	33,016	30,226	11,008	42,379	28,461	25,109	22,932	15,522	5,006	5,058	6	0
382	17,715	4,560	6,900	33,140	13,052	41,876	11,848	10,106	10,280	33,844	30,820	11,334	43,205	29,231	25,960	23,600	15,979	5,152	5,184	0	1
383	8,262	1,848	3,026	14,081	5,990	18,965	5,334	4,336	4,319	15,536	13,636	4,735	19,913	12,493	11,432	9,987	6,790	2,420	2,167	108	-1
384	9,072	2,090	3,267	15,115	6,576	20,043	5,976	4,855	4,819	16,607	14,690	5,318	21,517	13,889	12,535	10,855	7,374	2,521	2,463	100	-1
385	12,447	2,724	4,159	23,464	7,882	28,182	7,334	6,081	6,513	21,491	19,341	7,525	29,394	19,717	16,241	14,382	9,854	3,283	3,306	68	-1
386	19,125	4,506	6,830	35,875	13,027	45,582	11,858	9,799	10,316	34,376	31,113	12,124	46,628	31,085	26,173	23,467	16,080	5,347	5,321	0	1
387	18,645	4,382	6,601	34,859	12,676	44,513	11,372	9,511	10,040	33,582	30,345	11,740	45,438	30,177	25,519	22,832	15,656	5,233	5,137	4	0
388	1,228	0,217	0,344	2,436	0,626	2,577	0,630	0,548	0,592	1,963	1,562	0,689	2,904	1,769	1,543	1,201	0,847	0,346	0,277	172	-1
389	1,065	0,160	0,258	2,052	0,479	2,085	0,512	0,414	0,433	1,478	1,191	0,560	2,308	1,503	1,248	0,850	0,608	0,293	0,213	176	-1
390	13,304	3,056	4,539	24,894	8,726	30,679	7,941	6,586	7,034	22,974	20,930	8,243	31,916	21,446	17,547	15,511	10,655	3,494	3,618	60	-1
391	9,507	1,928	2,906	17,911	5,543	20,976	5,257	4,327	4,731	16,149	14,216	5,325	20,969	14,747	12,025	10,545	7,181	2,542	2,307	98	-1
392	18,378	4,346	6,480	34,254	12,458	43,770	11,242	9,319	9,825	32,885	29,841	11,542	44,899	29,719	25,010	22,356	15,283	5,160	5,044	6	0
393	17,893	4,778	6,666	30,665	13,769	43,044	10,886	9,509	10,356	34,969	31,444	11,085	45,740	27,350	26,256	23,064	15,372	4,902	5,416	0	1
394	15,262	4,070	5,791	25,451	12,084	37,145	9,378	8,168	8,957	29,551	27,365	9,718	38,941	23,271	22,903	19,898	13,249	4,141	4,755	22	-1
395	3,677	0,950	1,324	6,000	2,733	8,544	2,332	1,826	1,982	7,169	6,077	2,132	9,636	5,199	5,114	4,570	3,036	0,962	1,084	136	-1
396	7,272	2,015	2,657	11,938	5,668	16,903	4,668	3,868	4,267	13,958	11,935	4,655	18,956	10,919	10,644	9,254	6,243	1,907	2,252	102	-1
397	8,410	2,339	3,077	14,175	6,474	19,844	5,266	4,516	4,950	16,143	14,458	5,364	21,897	12,751	12,322	10,799	7,261	2,224	2,585	82	-1
398	11,585	3,052	4,265	18,580	9,199	28,015	6,954	6,014	6,700	21,938	20,541	7,340	28,184	17,630	17,302	15,060	9,972	2,992	3,594	60	-1
399	17,003	4,489	6,465	29,081	13,272	41,090	10,408	9,136	10,000	33,573	30,140	10,630	43,980	26,010	25,182	22,142	14,783	4,658	5,232	6	0
400	16,840	4,446	6,384	28,689	13,135	40,601	10,321	9,031	9,890	33,203	29,833	10,512	43,503	25,670	24,893	21,917	14,614	4,608	5,172	10	0

Results of testing decision trees and results of testing of voting procedures

No	Status	Dec tree (4.8)	Dec tree (4.9)	Dec tree (4.10)	Voting "2 of 3"	Voting "at least 1"		No	Status	Dec tree (4.8)	Dec tree (4.9)	Dec tree (4.10)	Voting "2 of 3"	Voting "at least 1"
1	1	0	0	0	0	0		39	0	0	0	0	0	0
2	0	0	0	0	0	0		40	0	0	0	0	0	0
3	0	0	0	0	0	0		41	1	1	1	1	1	1
4	0	0	0	0	0	0		42	0	0	0	0	0	0
5	1	1	1	1	1	1		43	1	1	1	1	1	1
6	1	1	1	0	1	1		44	0	0	1	0	0	1
7	0	0	0	0	0	0		45	1	1	1	1	1	1
8	0	0	0	0	0	0		46	0	0	0	0	0	0
9	1	1	1	1	1	1		47	0	0	0	0	0	0
10	0	0	0	0	0	0		48	0	0	0	0	0	0
11	0	0	0	0	0	0		49	1	1	1	1	1	1
12	0	0	0	0	0	0		50	0	0	0	0	0	0
13	0	0	0	0	0	0		51	1	0	1	0	0	1
14	1	1	1	1	1	1		52	0	0	0	0	0	0
15	1	1	0	1	1	1		53	0	0	0	0	0	0
16	0	1	1	1	1	1		54	1	1	1	1	1	1
17	0	0	0	0	0	0		55	0	0	1	0	0	1
18	0	1	0	0	0	1		56	0	0	0	0	0	0
19	0	0	0	0	0	0		57	0	0	0	0	0	0
20	1	1	1	1	1	1		58	1	1	1	1	1	1
21	1	1	0	1	1	1		59	0	1	1	1	1	1
22	0	0	0	0	0	0		60	0	0	0	0	0	0
23	0	0	0	0	0	0		61	1	1	1	1	1	1
24	1	1	1	1	1	1		62	0	0	0	0	0	0
25	0	1	1	0	1	1		63	1	1	1	1	1	1
26	0	0	0	0	0	0		64	0	0	0	0	0	0
27	1	1	1	1	1	1		65	0	1	1	1	1	1
28	0	0	0	0	0	0		66	0	0	0	0	0	0
29	0	0	0	0	0	0		67	0	0	0	0	0	0
30	0	0	0	0	0	0		68	1	1	1	1	1	1
31	0	0	0	0	0	0		69	1	1	1	1	1	1
32	1	0	0	0	0	0		70	0	0	0	0	0	0
33	1	1	1	1	1	1		71	0	0	1	0	0	1
34	0	1	1	1	1	1		72	1	1	1	1	1	1
35	0	1	1	1	1	1		73	0	0	0	0	0	0
36	0	0	0	0	0	0		74	0	0	0	0	0	0
37	0	0	0	0	0	0		75	0	0	0	0	0	0
38	1	1	1	1	1	1								

Appendix A3. Algebraic Bayes' Network related equations

A3.1. Equations over probabilities of logic formulae for knowledge pieces of rank 2 and 3 used to process experts' information to design locally consistent ABN

Knowledge piece of rank 2

$E_1^{(2)}\ p(x_1)+p(\neg x_1)=1,$

$E_2^{(2)}:\ p(x_2)+p(\neg x_2)=1,$

$E_3^{(2)}:\ p(x_1)+p(x_2)-p(x_1x_2)+p(\neg x_1\neg x_2)=1,$

$E_4^{(2)}:\ p(x_2)-p(x_1x_2)-p(\neg x_1x_2)=0,$

$E_5^{(2)}:\ p(x_1)-p(x_1x_2)-p(x_1\neg x_2)=0,$

$E_6^{(2)}:\ p(x_1\vee x_2)=p(x_1)+p(x_2)-p(x_1x_2),$

$E_7^{(2)}:\ p(\neg x_1\vee x_2)=1-p(x_1)+p(x_1x_2),$

$E_8^{(2)}:\ p(x_1\vee\neg x_2)=1-p(x_2)+p(x_1x_2),$

$E_9^{(2)}:\ p(\neg x_1\vee\neg x_2)=1-p(x_1x_2).$

Thus, in the two- proposition case only three unknowns have to be calculated. To transform experts' data to the ABN structure it is enough to select as unknowns the following ones: $p(x_1)$, $p(x_2)$, $p(x_2x_2)$.

Knowledge piece of rank 3

$E_1^{(3)}:\ p(x_1)+p(\neg x_1)=1,$

$E_2^{(3)}:\ p(x_2)+p(\neg x_2)=1,$

$E_3^{(3)}:\ p(x_3)+p(\neg x_3)=1,$

$E_4^{(3)}:\ p(x_1)+p(x_2)-p(x_1x_2)+p(\neg x_1\neg x_2)=1,$

$E_5^{(3)}:\ p(x_1)+p(x_3)-p(x_1x_3)+p(\neg x_1\neg x_3)=1,$

$E_6^{(3)}:\ p(x_2)+p(x_3)-p(x_2x_3)+p(\neg x_2\neg x_3)=1,$

$$E_7^{(3)}:\ p(x_1)+p(x_2)+p(x_3)-p(x_1x_2)-p(x_1x_3)-p(x_2x_3)+p(x_1x_2x_3)+ \\ +p(\neg x_1\neg x_2\neg x_3)=1,$$

$E_8^{(3)}:\ p(x_2)-p(x_1x_2)-p(\neg x_1x_2)=0,$

$E_9^{(3)}:\ p(x_1)-p(x_1x_2)-p(x_1\neg x_2)=0,$

$E_{10}^{(3)}:\ p(x_3)-p(x_1x_3)-p(\neg x_1x_3)=0,$

$E_{11}^{(3)}:\ p(x_1)-p(x_1x_3)-p(x_1\neg x_3)=0,$

$E_{12}^{(3)}:\ p(x_3)-p(x_2x_3)-p(\neg x_2x_3)=0,$

$E_{13}^{(3)}:\ p(x_2)-p(x_2x_3)-p(x_1\neg x_3)=0,$

$E_{14}^{(3)}:\ p(x_2x_3)-p(x_1x_2x_3)-p(\neg x_1x_2x_3)=0,$

$E_{15}^{(3)}:\ p(x_1x_3)-p(x_1x_2x_3)-p(x_1\neg x_2x_3)=0,$

$E_{16}^{(3)}:\ p(x_1x_2)-p(x_1x_2x_3)-p(x_1x_2\neg x_3)=0,$

$E_{17}^{(3)}:\ p(x_3)-p(x_1x_3)-p(x_2x_3)+p(x_1x_2x_3)-p(\neg x_1\neg x_2x_3)=0,$

$E_{18}^{(3)}:\ p(x_2)-p(x_1x_2)-p(x_2x_3)+p(x_1x_2x_3)-p(\neg x_1x_2\neg x_3)=0,$

$$E_{19}^{(3)}: p(x_1)-p(x_1x_2)-p(x_1x_3)+p(x_1x_2x_3)-p(x_1\neg x_2\neg x_3)=0,$$

$$E_{20}^{(3)}: p(x_1 \vee x_2)=p(x_1)+p(x_2)-p(x_1x_2),$$

$$E_{21}^{(3)}: p(\neg x_1 \vee x_2)=1-p(x_1)+p(x_1x_2),$$

$$E_{22}^{(3)}: p(x_1 \vee \neg x_2)=1-p(x_2)+p(x_1x_2),$$

$$E_{23}^{(3)}: p(\neg x_1 \vee \neg x_2)=1-p(x_1x_2),$$

$$E_{24}^{(3)}: p(x_1 \vee x_3)=p(x_1)+p(x_3)-p(x_1x_3),$$

$$E_{25}^{(3)}: p(\neg x_1 \vee x_3)=1-p(x_1)+p(x_1x_3),$$

$$E_{26}^{(3)}: p(x_1 \vee \neg x_3)=1-p(x_3)+p(x_1x_3),$$

$$E_{27}^{(3)}: p(\neg x_1 \vee \neg x_3)=1-p(x_1x_3),$$

$$E_{28}^{(3)}: p(x_2 \vee x_3)=p(x_2)+p(x_3)-p(x_2x_3),$$

$$E_{29}^{(3)}: p(\neg x_2 \vee x_3)=1-p(x_2)+p(x_2x_3),$$

$$E_{30}^{(3)}: p(x_2 \vee \neg x_3)=1-p(x_3)+p(x_2x_3),$$

$$E_{31}^{(3)}: p(\neg x_2 \vee \neg x_3)=1-p(x_2x_3),$$

$$\begin{aligned} E_{32}^{(3)}: p(x_1 \vee x_2 \vee x_3)&=p(x_1)+p(x_2)+p(x_3)- \\ &-p(x_1x_2)-p(x_1x_3)-p(x_2x_3)+p(x_1x_2x_3), \end{aligned}$$

$$E_{33}^{(3)}: p(\neg x_1 \vee x_2 \vee x_3)=1-p(x_1)+p(x_1x_2)+p(x_1x_3)-p(x_1x_2x_3),$$

$$E_{34}^{(3)}: p(x_1 \vee \neg x_2 \vee x_3)=1-p(x_2)+p(x_1x_2)+p(x_2x_3)-p(x_1x_2x_3),$$

$$E_{35}^{(3)}: p(x_1 \vee x_2 \vee \neg x_3)=1-p(x_3)+p(x_1x_3)+p(x_2x_3)-p(x_1x_2x_3),$$

$$E_{36}^{(3)}: p(\neg x_1 \vee \neg x_2 \vee x_3)=1-p(x_1x_2)+p(x_1x_2x_3),$$

$$E_{37}^{(3)}: p(\neg x_1 \vee x_2 \vee \neg x_3)=1-p(x_1x_3)+p(x_1x_2x_3),$$

$$E_{38}^{(3)}: p(x_1 \vee \neg x_2 \vee \neg x_3)=1-p(x_2x_3)+p(x_1x_2x_3),$$

$$E_{39}^{(3)}: p(\neg x_1 \vee \neg x_2 \vee \neg x_3)=1-p(x_1x_2x_3).$$

Thus, in the three–proposition case only seven unknowns have to be calculated. To transform experts' data to the ABN structure it is enough to select as unknowns the following ones: $p(x_1)$, $p(x_2)$, $p(x_3)$, $p(x_1x_2)$, $p(x_1x_3)$, $p(x_2x_3)$ and $p(x_1x_2x_3)$.

A3.2. Conditions of Consistency of knowledge Pieces of rank 2, 3 and 4 expressed in terms of positive conjunctions

Conditions of Consistency of knowledge piece of rank 2

$$C_1^{(2)}\ p(x_1) \le 1,$$

$$C_2^{(2)}: p(x_2) \le 1,$$

$$C_3^{(2)}: p(x_1)+p(x_2)-p(x_1x_2) \le 1,$$

$$C_4^{(2)}: p(x_2)-p(x_1x_2) \ge 0,$$

$$C_5^{(2)}: p(x_1)-p(x_1x_2) \ge 0,$$

$$p(x_i) \ge 0,\ i=1,2,\ p(x_1x_2) \ge 0.$$

Conditions of Consistency of knowledge piece of rank 3

$$C_1^{(3)}: p(x_1) \le 1,$$

$C_2^{(3)}:\ p(x_2)\le 1,$

$C_3^{(3)}:\ p(x_3)\le 1,$

$C_4^{(3)}:\ p(x_1)+p(x_2)-p(x_1x_2)\le 1,$

$C_5^{(3)}:\ p(x_1)+p(x_3)-p(x_1x_3)\le 1,$

$C_6^{(3)}:\ p(x_2)+p(x_3)-p(x_2x_3)\le 1,$

$C_7^{(3)}:\ p(x_1)+p(x_2)+p(x_3)-p(x_1x_2)-p(x_1x_3)-p(x_2x_3)+p(x_1x_2x_3)\le 1,$

$C_8^{(3)}:\ p(x_2)-p(x_1x_2)\ge 0,$

$C_9^{(3)}:\ p(x_1)-p(x_1x_2)\ge 0,$

$C_{10}^{(3)}:\ p(x_3)-p(x_1x_3)\ge 0,$

$C_{11}^{(3)}:\ p(x_1)-p(x_1x_3)\ge 0,$

$C_{12}^{(3)}:\ p(x_3)-p(x_2x_3)\ge 0,$

$C_{13}^{(3)}:\ p(x_2)-p(x_2x_3)\ge 0,$

$C_{14}^{(3)}:\ p(x_2x_3)-p(x_1x_2x_3)\ge 0,$

$C_{15}^{(3)}:\ p(x_1x_3)-p(x_1x_2x_3)\ge 0,$

$C_{16}^{(3)}:\ p(x_1x_2)-p(x_1x_2x_3)\ge 0,$

$C_{17}^{(3)}:\ p(x_3)-p(x_1x_3)-p(x_2x_3)+p(x_1x_2x_3)\ge 0,$

$C_{18}^{(3)}:\ p(x_2)-p(x_1x_2)-p(x_2x_3)+p(x_1x_2x_3)\ge 0,$

$C_{19}^{(3)}:\ p(x_1)-p(x_1x_2)-p(x_1x_3)+p(x_1x_2x_3)\ge 0,$

$p(x_i)\ge 0,\ i=1,2,3,\ p(x_ix_j)\ge 0,\ i,j=1,2,3,\ i>j,\ p(x_1x_2x_3)\ge 0.$

Conditions of Consistency of knowledge piece of rank 4

$C_1^{(4)}:\ p(x_1)\le 1,$

$C_2^{(4)}:\ p(x_2)\le 1,$

$C_3^{(4)}:\ p(x_3)\le 1,$

$C_4^{(4)}:\ p(x_4)\le 1,$

$C_5^{(4)}:\ p(x_1)+p(x_2)-p(x_1x_2)\le 1,$

$C_6^{(4)}:\ p(x_1)+p(x_3)-p(x_1x_3)\le 1,$

$C_7^{(4)}:\ p(x_1)+p(x_4)-p(x_1x_4)\le 1,$

$C_8^{(4)}:\ p(x_2)+p(x_3)-p(x_2x_3)\le 1,$

$C_9^{(4)}:\ p(x_2)+p(x_4)-p(x_2x_4)\le 1,$

$C_{10}^{(4)}:\ p(x_3)+p(x_4)-p(x_3x_4)\le 1,$

$C_{11}^{(4)}:\ p(x_1)+p(x_2)+p(x_3)-p(x_1x_2)-p(x_1x_3)-p(x_2x_3)+p(x_1x_2x_3)\le 1,$

$C_{12}^{(4)}:\ p(x_1)+p(x_2)+p(x_4)-p(x_1x_2)-p(x_1x_4)-p(x_2x_4)+p(x_1x_2x_4)\le 1,$

$C_{13}^{(4)}:\ p(x_1)+p(x_3)+p(x_4)-p(x_1x_3)-p(x_1x_4)-p(x_3x_4)+p(x_1x_3x_4)\le 1,$

$C_{14}^{(4)}:\ p(x_2)+p(x_3)+p(x_4)-p(x_2x_3)-p(x_2x_4)-p(x_3x_4)+p(x_2x_3x_4)\le 1,$

Appendix A3. Algebraic Bayes' network related equations

$$\begin{aligned} C_{15}^{(4)}:\ & p(x_1)+p(x_2)+p(x_3)+p(x_4)-p(x_1x_2)-p(x_1x_3)-p(x_1x_4)-p(x_2x_3)- \\ & -p(x_2x_4)-p(x_3x_4)+p(x_1x_2x_3)+p(x_1x_2x_4)+p(x_1x_3x_4)+p(x_2x_3x_4)- \\ & -p(x_1x_2x_3x_4)\leq 1, \end{aligned}$$

$C_{16}^{(4)}:\ p(x_2)-p(x_1x_2)\geq 0,$
$C_{17}^{(4)}:\ p(x_1)-p(x_1x_2)\geq 0,$
$C_{18}^{(4)}:\ p(x_3)-p(x_1x_3)\geq 0,$
$C_{19}^{(4)}:\ p(x_1)-p(x_1x_3)\geq 0,$
$C_{20}^{(4)}:\ p(x_4)-p(x_1x_4)\geq 0,$
$C_{21}^{(4)}:\ p(x_1)-p(x_1x_4)\geq 0,$
$C_{22}^{(4)}:\ p(x_3)-p(x_2x_3)\geq 0,$
$C_{23}^{(4)}:\ p(x_2)-p(x_2x_3)\geq 0,$
$C_{24}^{(4)}:\ p(x_4)-p(x_2x_4)\geq 0,$
$C_{25}^{(4)}:\ p(x_2)-p(x_2x_4)\geq 0,$
$C_{26}^{(4)}:\ p(x_4)-p(x_3x_4)\geq 0,$
$C_{27}^{(4)}:\ p(x_3)-p(x_3x_4)\geq 0,$

$C_{28}^{(4)}:\ p(x_2x_3)-p(x_1x_2x_3)\geq 0,$
$C_{29}^{(4)}:\ p(x_1x_3)-p(x_1x_2x_3)\geq 0,$
$C_{30}^{(4)}:\ p(x_1x_2)-p(x_1x_2x_3)\geq 0,$

$C_{31}^{(4)}:\ p(x_2x_4)-p(x_1x_2x_4)\geq 0,$
$C_{32}^{(4)}:\ p(x_1x_4)-p(x_1x_2x_4)\geq 0,$
$C_{33}^{(4)}:\ p(x_1x_2)-p(x_1x_2x_4)\geq 0,$

$C_{34}^{(4)}:\ p(x_3x_4)-p(x_1x_3x_4)\geq 0,$
$C_{35}^{(4)}:\ p(x_1x_4)-p(x_1x_3x_4)\geq 0,$
$C_{36}^{(4)}:\ p(x_1x_3)-p(x_1x_3x_4)\geq 0,$

$C_{37}^{(4)}:\ p(x_3x_4)-p(x_2x_3x_4)\geq 0,$
$C_{38}^{(4)}:\ p(x_2x_4)-p(x_2x_3x_4)\geq 0,$
$C_{39}^{(4)}:\ p(x_2x_3)-p(x_2x_3x_4)\geq 0,$

$C_{40}^{(4}:\ p(x_3)-p(x_1x_3)-p(x_2x_3)+p(x_1x_2x_3)\geq 0,$
$C_{41}^{(4)}:\ p(x_2)-p(x_1x_2)-p(x_2x_3)+p(x_1x_2x_3)\geq 0,$
$C_{42}^{(4)}:\ p(x_1)-p(x_1x_2)-p(x_1x_3)+p(x_1x_2x_3)\geq 0,$

$C_{43}^{(4)}:\ p(x_4)-p(x_1x_4)-p(x_2x_4)+p(x_1x_2x_4)\geq 0,$
$C_{44}^{(4)}:\ p(x_2)-p(x_1x_2)-p(x_2x_4)+p(x_1x_2x_4)\geq 0,$
$C_{45}^{(4)}:\ p(x_1)-p(x_1x_2)-p(x_1x_4)+p(x_1x_2x_4)\geq 0,$

$C_{46}^{(4}:\ p(x_4)-p(x_1x_4)-p(x_3x_4)+p(x_1x_3x_4)\geq 0,$
$C_{47}^{(4)}:\ p(x_3)-p(x_1x_3)-p(x_3x_4)+p(x_1x_3x_4)\geq 0,$

Appendix A3. Algebraic Bayes' network related equations

$C_{48}{}^{(4)}: p(x_1)-p(x_1x_3)-p(x_1x_4)+p(x_1x_3x_4)\geq 0,$

$C_{49}{}^{(4)}: p(x_4)-p(x_2x_4)-p(x_3x_4)+p(x_2x_3x_4)\geq 0,$

$C_{50}{}^{(4)}: p(x_3)-p(x_2x_3)-p(x_3x_4)+p(x_2x_3x_4)\geq 0,$

$C_{51}{}^{(4)}: p(x_2)-p(x_2x_3)-p(x_2x_4)+p(x_2x_3x_4)\geq 0,$

$p(x_i)\geq 0,\ i=1,2,3,4,\ p(x_ix_j)\geq 0,\ i,j=1,2,3,4,\ i>j,$

$p(x_ix_jx_k)\geq 0,\ i,j,k=1,2,3,4,\ i>j>k,\ p(x_1x_2x_3x_4)\geq 0,$

$C_{52}{}^{(4)}: p(x_1x_2x_3)-p(x_1x_2x_3x_4)\geq 0,$

$C_{53}{}^{(4)}: p(x_1x_2x_4)-p(x_1x_2x_3x_4)\geq 0,$

$C_{54}{}^{(4)}: p(x_1x_3x_4)-p(x_1x_2x_3x_4)\geq 0,$

$C_{55}{}^{(4)}: p(x_2x_3x_4)-p(x_1x_2x_3x_4)\geq 0,$

$C_{56}{}^{(4)}: p(x_1x_2)-p(x_1x_2x_3)-p(x_1x_2x_4)+p(x_1x_2x_3x_4)\geq 0,$

$C_{57}{}^{(4)}: p(x_1x_3)-p(x_1x_2x_3)-p(x_1x_2x_4)+p(x_1x_2x_3x_4)\geq 0,$

$C_{58}{}^{(4)}: p(x_1x_4)-p(x_1x_2x_4)-p(x_1x_3x_4)+p(x_1x_2x_3x_4)\geq 0,$

$C_{59}{}^{(4)}: p(x_2x_3)-p(x_1x_2x_3)-p(x_2x_3x_4)+p(x_1x_2x_3x_4)\geq 0,$

$C_{60}{}^{(4)}: p(x_2x_4)-p(x_1x_2x_4)-p(x_2x_3x_4)+p(x_1x_2x_3x_4)\geq 0,$

$C_{61}{}^{(4)}: p(x_3x_4)-p(x_1x_3x_4)-p(x_2x_3x_4)+p(x_1x_2x_3x_4)\geq 0,$

$$C_{62}{}^{(4)}: p(x_1)-p(x_1x_2)-p(x_1x_3)-p(x_1x_4)+p(x_1x_2x_3)+p(x_1x_2x_4)+ \\ +p(x_1x_3x_4)-p(x_1x_2x_3x_4)\geq 0,$$

$$C_{63}{}^{(4)}: p(x_2)-p(x_1x_2)-p(x_2x_3)-p(x_2x_4)+p(x_1x_2x_3)+p(x_1x_2x_4)+ \\ +p(x_2x_3x_4)-p(x_1x_2x_3x_4)\geq 0,$$

$$C_{64}{}^{(4)}: p(x_3)-p(x_1x_3)-p(x_2x_3)-p(x_3x_4)+p(x_1x_2x_3)+p(x_1x_3x_4)+ \\ +p(x_2x_3x_4)-p(x_1x_2x_3x_4)\geq 0,$$

$$C_{65}{}^{(4)}: p(x_4)-p(x_1x_4)-p(x_2x_4)-p(x_3x_4)+p(x_1x_2x_4)+p(x_1x_3x_4)+ \\ +p(x_2x_3x_4)-p(x_1x_2x_3x_4)\geq 0,$$

$p(x_1x_2x_3x_4)\geq 0.$